The Dissertation Committee for Rajiv Soundararajan
certifies that this is the approved version of the following dissertation:

Information Theoretic Methods in Distributed Compression and Visual Quality Assessment

Committee:

Alan C. Bovik, Supervisor

Sriram Vishwanath, Supervisor

Wilson Geisler

Joydeep Ghosh

Gustavo de Veciana

Haris Vikalo

Information Theoretic Methods in Distributed Compression and Visual Quality Assessment

by

Rajiv Soundararajan, B. E.; M. S. E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2012

Dedicated to my parents

and grandmother

Acknowledgments

My PhD life has provided me an extremely satisfying and enjoyable academic experience. I am indebted to my supervisors Al Bovik and Sriram Vishwanath for their inspiration, support and guidance. Dr. Bovik has been a pillar of strength, listening patiently to my everyday successes and failures, encouraging me and instilling hope at all times. I greatly benefited from his vision for research and constructive comments and consider myself fortunate to have worked with him. I am extremely grateful to Sriram for introducing me to research, helping develop my research thinking and improve my presentation skills and offering valuable advice at different points. He is an amazing source of research ideas and I learned a lot about generating such ideas from him. I also thoroughly enjoyed the wonderful opportunities that he gave me to collaborate with renowned researchers at other universities.

I would like to thank my other dissertation committee members: Wilson Geisler, Joydeep Ghosh, Gustavo de Veciana and Haris Vikalo for having served on my committee. Their comments and suggestions greatly helped improve different aspects of this dissertation. I thank Aaron Wagner at Cornell University, for his valuable guidance through a chapter of this dissertation, insightful discussions and helpful comments. I also thank Tsachy Weissman for hosting me during my visit to Stanford University, letting me participate in

his group meetings and enabling a fantastic research experience. I am thankful to Pierre Costa at AT&T Labs, Austin for having funded my research for many years and helping me gain useful industry experience. I thank Janet Preuss and the other staff at WNCG for having meticulously taken care of my appointments and reimbursements right through my PhD.

My graduate school life would be incomplete without the excellent interactions that I've had with fellow graduate students at UT Austin. I cherish all the graduate classes that I took with Jubin Jose and Aneesh Reddy who joined UT with me and the numerous research discussions that I've indulged in with them. I treasure the various conversations I've had with fellow LINC group members, Brian Smith, Chulhan Lee, Caleb Lo, Sriram Sridharan, Shreeshankar Bodas, Amin Jafarian, Shweta Agrawal, Sang Hyun Lee, Goochul Chung, Kumar Appaiah, Abhik Das, Sharayu Moharir, Ioannis Mitliagkas, Deepjyoti Deka, Ankit Rawat, Hongbo Si, Cong Li, Avhishek Chatterjee and many others through group meetings and otherwise. I also thank them for helping me in my presentations and research over the years. I had amazing times collaborating on various papers with Kalpana Seshadrinathan, Anush Moorthy, Ozan Koyluoglu, Anish Mittal, Gautam Muralidhar and Dinesh Jayaraman. I have fond memories of visiting AT&T Labs with Sumohana Channappayya and Joonsoo Lee. I also thank my fellow LIVE lab members, Sina Jahanbin, Michele Saad, Ming Jun Chen, Tony Su and others for great times.

I had an awesome social life at Austin which kept me in great spirits

during the course of my PhD. I thank my room mates Adithya, Karthik and Sriram for having put up with me and for having been very adjusting and accommodating. I shall never forget the fabulous times with Sriram, Suudhan, Vivek and Yasas playing tennis, watching movies and engaging in long philosophical conversations. I also thank my friends from undergrad who have been a great source of warmth and happiness through all these years.

My parents have always stood by me and offered their unflinching support to my every endeavor in life. Indeed, this dissertation would not have been possible without their blessings and good wishes. I have no words to express my gratitude to them for their affection and encouragement at all times. My grandmother has been a great companion since my younger days and I thank her for her blessings. I also wish to thank my relatives and cousins for being wonderful well-wishers.

Information Theoretic Methods in Distributed Compression and Visual Quality Assessment

Publication No. _____

Rajiv Soundararajan, Ph.D.
The University of Texas at Austin, 2012

Supervisors: Alan C. Bovik
Sriram Vishwanath

Distributed compression and quality assessment (QA) are essential ingredients in the design and analysis of networked signal processing systems with voluminous data. Distributed source coding techniques enable the efficient utilization of available resources and are extremely important in a multitude of data intensive applications including image and video. The quality analysis of such systems is also equally important in providing benchmarks on performance leading to improved design and control. This dissertation approaches the complementary problems of distributed compression and quality assessment using information theoretic methods. While such an approach provides intuition on designing practical coding schemes for distributed compression, it directly yields image and video QA algorithms with excellent performance that can be employed in practice.

This dissertation considers the information theoretic study of sophisticated problems in distributed compression including, multiterminal multi-

ple description coding, multiterminal source coding through relays and joint source channel coding of correlated sources over wireless channels. Random and/or structured codes are developed and shown to be optimal or near optimal through novel bounds on performance. While lattices play an important role in designing near optimal codes for multiterminal source coding through relays and joint source channel coding over multiple access channels, time sharing random Gaussian codebooks is optimal for a wide range of system parameters in the multiterminal multiple description coding problem.

The dissertation also addresses the challenging problem of reduced reference image and video QA. A family of novel reduced reference image and video QA algorithms are developed based on spatial and temporal entropic differences. While the QA algorithms for still images only compute spatial entropic differences, the video QA algorithms compute both spatial and temporal entropic differences and combine them in a perceptually relevant manner. These algorithms attain excellent performances in terms of correlation with human judgments of quality on large QA databases. The framework developed also enables the study of the degradation in performance of QA algorithms from full reference information to almost no information from the reference image or video.

Table of Contents

List of Figures

Chapter 1

Introduction

In recent years, there has been an explosion in the amount of data that needs to be stored and communicated, particularly, visual data such as image and video. With the advent of the internet and mobile devices, understanding how data may be efficiently compressed and processed in networks is of significant interest. The problem of distributed compression precisely deals with this question of how correlated data may be compressed in a distributed fashion in networks, with or without distortions. The problem has applications in a variety of domains including sensor networks, video surveillance systems, media delivery over networks, distributed data storage and so on. Owing to the voluminous amount of data involved in all these applications, it is impossible to store or communicate such data without lossy compression. The challenge in such problems is to exploit the correlation in the underlying data at different nodes in a network in a distributed manner and design efficient coding schemes. The goal is to thereby achieve better performance than systems in which individual nodes merely compress their own data.

On the other hand, most data like image and video, are ultimately meant for human consumption. The impact of digital image and video in

today's life is pervasive with applications ranging from video streaming over mobile devices and computers to video conferencing, surveillance and digital cinema. This renders the question of user experience through quality monitoring and control extremely important. It is desirable to design such systems in a manner that enhances the visual experience of users. The field of image and video quality assessment (QA) seeks to partially address the question of how to quantitatively model user experience, and how to use these models to predict visual quality in accordance with visual perception. Image/video quality assessment is an essential step towards ensuring and meeting quality of service demands of various users. Additionally, the design of systems optimized for such quality indices could potentially help in significantly enhancing the visual experience. The challenge in this problem is that quality assessment is not merely a measurement of artifacts or distortions in visual signals, but measurement of those distortions that are perceivable by humans.

Compression and quality assessment are essentially complementary engineering optimization problems. In compression, given a distortion criterion, we seek an optimal compression algorithm, the objective of the optimization problem being the rate of compression. Conversely in quality assessment, given a distorted image, we seek an optimal quality assessment algorithm, the objective here being the correlation with human judgments of quality. This dissertation addresses sophisticated problems in both these domains using information theoretic tools as a common theme and can be divided into two parts: distributed compression and visual quality assessment. We provide a

summary of the research problems studied in each of these domains along with the main contributions of the dissertation in the following.

1.1 Distributed Compression

Point-to-point compression is a fairly well established area of research today, both in theory and practice, as demonstrated by Shannon's rate distortion theorem [15] and practical audio, image and video codecs such as MP3, JPEG and MPEG. The holy grail of lossless and lossy network compression is in understanding the fundamental limits and optimal coding strategies of systems with multiple sources, relays and destinations connected arbitrarily between them. We desire to gain an understanding of the *rate-distortion region* of this distributed compression problem, which is a highly non-trivial task for a large class of such problems. Our understanding of lossy compression schemes even in simple networks is quite limited. On the theoretical side, optimal coding schemes are known only for the Gaussian two terminal source coding problem (the simplest many-to-one network) [83] and the Gaussian multiple description problem (the simplest one-to-many network) [46, 85]. Nevertheless, the information theoretic study of such simple settings has enabled us gain some intuition on the underlying architecture of good coding schemes for practical systems [99]. Such schemes are shown to possess better performance than treating network compression problems as individual point-to-point problems. This motivates the information theoretic study of network compression problems in more general network settings, which is the focus of

this dissertation. As stated earlier, the goal of such a study is to learn which coding architectures are information theoretically optimal.

Since arbitrary networks have been difficult to analyze so far, network compression has only been studied primarily as many-to-one systems [45, 51, 79, 83] or one-to-many [13, 46, 85] systems. We take an important step towards understanding more general networks by analyzing many-to-many systems, in the form of multiterminal multiple description problems. We would like to know if it is sub-optimal to treat many-to-many systems as simpler networks and if so, what the optimal coding schemes in such scenarios are. We introduce a multiterminal multiple description problem called the Vacationing chief executive officer (CEO) problem. The vacationing- CEO problem combines the salient features of the so-called CEO problem [79] and the multiple-descriptions (MD) problem [13]. In this setting, noisy versions of a source are observed by two encoders, as in the CEO problem. In addition, we require that each encoder generate multiple descriptions of the source, as in the MD problem. The vacationing-CEO problem arises in asynchronous multicast networks, and solving it is an essential step in developing a general theory for multi-encoder and multi-decoder lossy compression. In this dissertation, an achievable sum rate and two sum rate lower bounds are presented for the quadratic Gaussian vacationing-CEO problem. These bounds exactly determine the optimal sum rate over a wide range of parameters.

Another step towards understanding compression in more general networks is the investigation of the role of relays. The key question in such

problems is in finding out how the relays should deal with incoming data in order to ensure that the reconstructions at nodes further down the network satisfy their distortion criteria while consuming minimum resources. Indeed, various researchers have studied the role of relays in distributed compression under different settings [16, 47, 78]. We study a multi-terminal source coding problem, where two terminals possess two (correlated) Gaussian sources to be compressed and delivered to the destination through an intermediate relay. Unlike the CEO and conventional two-terminal source coding problems as well as point-to-point relay source coding problem, lattices are found to play an important role in achieving "good" rates for this problem setting. Two achievable strategies - compute-and-forward and compress-and-forward are used to develop achievable rates for this problem setting. For the symmetric case, the inner and outer bounds developed are shown to be within 1/2 bits of each other.

We are also interested in how distributed compression may be performed in the presence of wireless channels. One such formulation is a joint source-channel coding problem where correlated sources are required to be communicated in a distributed manner over a multiple access channel, a problem in which very little progress has been possible over the years. A fundamental question in such problems is whether it is optimal to separate source coding from channel coding. We would also like to know if there are other schemes that could exploit the correlation between the sources for better transmission. While the solution for such a general joint source channel coding problem is un-

known, different aspects of the problem have been considered in [14, 25, 34, 42]. This dissertation considers the problem of transmitting linear functions of two correlated Gaussian sources over a two-user additive Gaussian noise multiple access channel (MAC). The goal is to recover this linear function within an average mean squared error distortion criterion. Each transmitter has access to only one of the two Gaussian sources and is limited by an average power constraint. In this work, a lattice coding scheme and two lower bounds on the achievable distortion are presented. The lattice scheme achieves within a constant of a distortion lower bound if the signal to noise ratio (SNR) is greater than a threshold. Further, for the difference of correlated Gaussian sources, uncoded transmission is shown to be worse in performance to lattice coding methods for correlation coefficients above a threshold.

1.2 Visual Quality Assessment

Image and video QA algorithms can be broadly classified into full reference (reference available or FR) and no reference (reference not available or NR) algorithms. The mean squared error (MSE) has been used as a quality metric for a very long time, owing to its simplicity, despite having a very poor correlation with human perception [88]. The last decade has seen significant progress in the field of objective full reference image/video QA algorithms. The structural similarity index (SSIM) [89], visual information fidelity (VIF) [63], visual signal to noise ratio (VSNR) [9] and just noticeable difference (JND) metrix [56], are examples of successful full reference algorithms which

have been shown to perform well in predicting the quality scores of human subjects. However, in a number of applications, it is impossible or extremely expensive for the reference image or video to made available for quality computation. The development of NR QA algorithms is crucial for quality assessment in large networks.

Progress on no reference QA, however, has been very slow. Indeed, the progress that has been possible, has been on account of relaxing the no reference assumption in various ways. One approach is to devise NR algorithms for a specific type of distortion only [91], [64], [38]. This approach can be refined by assuming that the distorted image is subjected to a set of possible distortions known *a priori*. Training based NR image QA techniques have resulted in algorithms that perform at least as well as mean squared error, which has the benefit of a reference image [40, 55]. Alternatively, partial information about the reference can be made available, which can be used along with the distorted image to predict quality. This paradigm is known as reduced reference (RR) QA, which may or may not require knowledge of the distortion type.

We study the problem of automatic RR image QA algorithms from the point of view of image information change. Such changes are measured between the reference image and natural image approximations of the distorted image. Algorithms that measure differences between the entropies of wavelet coefficients of reference and distorted images as perceived by humans are designed. The algorithms differ in the data on which the entropy difference

is calculated and on the amount of information from the reference that is required for quality computation, ranging from almost full information to almost no information from the reference. A special case of this are algorithms that require just a single number from the reference for quality assessment. The algorithms are shown to correlate very well with subjective quality scores as demonstrated on the LIVE Image Quality Assessment Database and Tampere Image Database. The performance degradation as the amount of information is reduced is also studied.

We also extend this framework to video quality assessment. We present a family of reduced reference video QA algorithms that utilize spatial and temporal entropic differences. We adopt a hybrid approach of combining statistical models and perceptual principles to design QA algorithms. A Gaussian scale mixture model for the wavelet coefficients of frames as well as frame differences is used to measure the amount of spatial and temporal information differences between the reference and distorted videos respectively. The spatial and temporal information differences are combined to obtain the spatio-temporal reduced reference entropic differences. The algorithms are flexible in terms of the amount of side information required from the reference which can range between a single scalar per frame and the entire reference information. The spatio-temporal entropic differences are shown to correlate quite well with human judgments of quality as demonstrated by experiments on the LIVE Video Quality Assessment Database.

1.3 Organization

The rest of the dissertation is organized as follows. Chapters 2, 3 and 4 deal with distributed compression while Chapters 5 and 6 concern image and video quality assessment. The vacationing-CEO problem is studied in Chapter 2, multiterminal source coding through relays in Chapter 3 and communicating linear functions of correlated Gaussian sources over a Gaussian multiple access channel in Chapter 4. The reduced reference entropic differencing indices for image quality assessment are described in Chapter 5 while Chapter 6 contains the spatio-temporal entropic differencing indices for video quality assessment. We conclude the dissertation in Chapter 7.

Chapter 2

Sum Rate of the Vacationing CEO Problem

2.1 Introduction

The vacationing-chief executive officer (CEO) problem is one where multiple encoders compress noisy versions of a single source in a distributed manner with each encoder producing multiple descriptions. While network compression has been studied either as a multi-encoder single-description problem as in multi-terminal source coding, or as a single-encoder multiple description problem, this chapter considers a multi-terminal multiple description problem. The two-terminal three-description version of the vacationing-CEO problem is depicted in Fig. 2.1.

We have a single source S, and two corrupted versions of the source, X_1 and X_2, are available at the two encoders in the system. The encoders wish to communicate information about S to a decoder, i.e., the CEO, which they accomplish by each sending a data packet at time 1 and another at time 2. The CEO may be on vacation during time 1, time 2, neither, or both, and she cannot receive data packets when she is vacationing. We assume that the CEO's vacation schedule is unknown to the encoders. If the CEO works during time 1 and vacations during time 2, she expects to reproduce the source

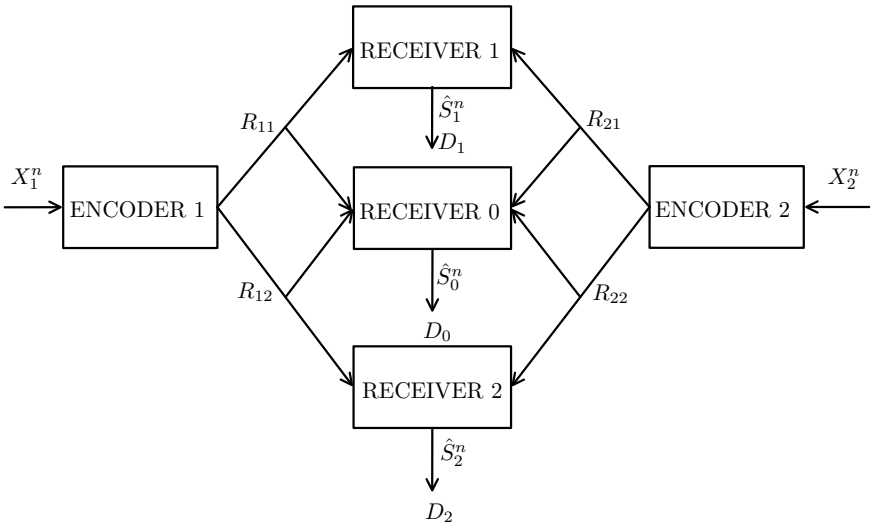


Figure 2.1: System model for the vacationing-CEO problem

S to distortion D_1. Likewise, if she works during time 2 and vacations during time 1, she expects to reproduce S to distortion D_2. If the CEO does not go on vacation and works during both periods, then she expects to reproduce S to distortion D_0. For convenience, we represent the three vacation states of the CEO by three separate receivers in Fig. 2.1. This problem generalizes both the CEO problem, by omitting the transmission at time 2, and the MD problem, by omitting X_2 and choosing X_1 to be an exact copy of S. But we shall see that solving the vacationing-CEO problem requires more than a direct combination of ideas from the CEO and MD problems.

The vacationing-CEO problem has applications in peer-to-peer networks where multiple encoders have incomplete or imperfect versions of the same source, which is to be compressed and shared with different destinations. Multiple descriptions are desired to alleviate the effects of packet loss. They

are also useful in wireless multicast scenarios in which receivers enter and depart the system unbeknown to the encoders. In this context we would like receivers to be able to arrive at any time and for their distortion to decrease the longer they remain in the system. The problem in Fig. 2.1 is a simple version of the scenario in which there are only two time steps. See [2–4] for additional discussion of this connection.

The following results on the vacationing-CEO problem are the contributions of this chapter of the dissertation:

1. We present two lower bounds on the sum rate of the vacation-CEO problem. The lower bounds combine converse techniques developed individually for the MD problem [46, 85] and the CEO problem [45, 87] in different ways. In fact, it is interesting to note that one of our lower bounds requires the use of *both* existing converse techniques for the CEO problem, as neither alone seems to be sufficient.

2. We develop an achievable scheme that generalizes the Berger-Tung scheme for multi-terminal source coding and the El Gamal-Cover scheme for multiple descriptions.

3. We show that a Gaussian scheme is optimal in the *low distortion* regime, i.e. when the distortion constraints at the individual receivers are 'close' to the distortion constraints at the central receiver. We also show that a time-sharing scheme with Gaussian codebooks is optimal in the *high distortion* regime.

4. We establish necessary and sufficient conditions for the achievability of no excess rate. That is, if the sum rate of the system is forced to be the minimal sum rate needed to meet the distortion constraint of Receiver 0, we establish the necessary conditions in order for the distortion constraints of the other two receivers to be achievable.

The rest of this chapter is organized as follows. The next subsection summarizes some of the key results in the field. In Section 2.2, we state the problem and state our main results. In Section 2.3, we describe the achievable strategy, and in Section 2.4, we prove the sum rate lower bounds. In Section 2.5 we prove that the achievable sum rate coincides with the first lower bound in the low-distortion regime, and in Section 2.6 we prove that the the achievable sum rate coincides with the second lower bound in the high-distortion regime.

2.1.1 Background

A single representation for a single source is today a fairly well established field of research [15]. When multiple representations and/or sources are involved, there are only a limited set of exact results known. The lossless compression of correlated sources, studied in [69] by Slepian and Wolf, is one of the early success stories in this domain. The problem of lossy compression with correlated side information at the decoder was considered in [97]. Subsequently, an achievable rate region for the multiple description (MD) problem was given by [13], which was shown to be optimal for the Gaussian two-terminal case in [46].

More recently, many new results have emerged in the field of Gaussian multiterminal source coding. In particular, the Gaussian CEO problem was introduced in [79] and solved in [45] and [51], where first the sum rate and later the entire rate region were characterized. While Gaussian schemes are shown to be optimal, the converse techniques primarily involve the entropy power inequality and the technique of formulating the outer bound as an optimization problem over suitably defined code parameters. [87] provides a simplified converse argument for the sum rate. The rate region of the Gaussian two-encoder problem was characterized in [83]. Lattice codes were shown to possess superior performance to random Gaussian codes for the distributed computation of certain linear functions of correlated Gaussian sources [33], [82]. On the multiple descriptions front, the sum rate of the Gaussian vector multiple description problem with any number of descriptions but distortion constraints only at individual and central receivers was established in [85] (see also [10, 86]). The rate distortion region for the multiple description problem involving discrete memoryless sources in the case of 'no excess' rate was solved in [1].

Researchers have also studied multi-source multi-desination compression problems in the context of "robust distributed source coding" [11]. In this setting, two corrupted versions of an underlying source are observed by two separate encoders, Encoder 1 and Encoder 2. Each encoder generates a single description of its observation. There are three receivers in the problem set up, Receivers 1, 2 and 3. Receiver 1 and Receiver 2 obtain the single description

from Encoder 1 and Encoder 2 respectively, while Receiver 3 obtains the descriptions from both encoders. Each receiver has its own distortion criteria and the objective of the problem is to find the smallest rates at which the descriptions at each encoder need to be generated satisfying all the distortion constraints.

As opposed to our problem formulation where each encoder generates two descriptions, the problem in [11] requires each encoder to generate only a single description. In fact, the vacationing CEO problem can be related to the problem in [11] when we constrain the rates $R_{12} = R_{21} = 0$. Although our setup is more general, we actually obtain more conclusive results than those in [19]. This is accomplished by focusing on the tradeoff between the sum rate $R_{11} + R_{12} + R_{21} + R_{22}$ and the triple of distortions D_0, D_1, and D_2, which provides a certain symmetry between the encoders and the descriptions. While the bounding technique of one of our two sum rate lower bounds is similar to the bounding technique used in [11], our other lower bound is established using a different technique that involves introducing two new auxiliary random variables, one for the observation of each encoder. Determining the full rate-distortion region for our problem would require determining the rate-distortion region for the problem in [11] and seems to be quite challenging.

2.2 Problem Statement and Main Result

We use capital letters to denote random variables and $\mathbb{E}[S]$ to denote the expected value of a random variable S. All logarithms used in the chapter are

natural logarithms. $\mathrm{Var}(S|T)$ denotes $\mathbb{E}_{S,T}[(S - \mathbb{E}[S|T])^2]$. For $\bar{S} = (S_1, S_2)$, $\mathrm{Cov}(\bar{S}|T)$ denotes the matrix $\mathbb{E}_{\bar{S},T}[(\bar{S} - \mathbb{E}[\bar{S}|T])(\bar{S} - \mathbb{E}[\bar{S}|T])^\dagger]$, where $\bar{S}^\dagger$ is the transpose of the vector $\bar{S}$.

2.2.1 Problem Statement

Let $\{X_{1i}\}_{i=1}^n$ and $\{X_{2i}\}_{i=1}^n$ be noisy observations of an underlying Gaussian source $\{S_i\}_{i=1}^n$, observed by two different encoders. The observations and the source are assumed to be independent and identically distributed (i.i.d.) over i. For each time instant i, the observations are given by

$$X_{1i} = S_i + N_{1i}$$

$$X_{2i} = S_i + N_{2i}$$

where N_{1i} and N_{2i} are Gaussian distributed with mean zero and variance $\sigma^2_{N_1}$ and $\sigma^2_{N_2}$. S_i has mean zero and variance σ^2_S. Encoder k observes $\{X_{ki}\}_{i=1}^n$ for $k = 1, 2$ and sends two descriptions given by $C_{kl} = f_{kl}(X_k^n)$, for $l = 1, 2$ to two receivers. Let R_{kl} be the rate of transmission from Encoder k to Receiver l. Receiver l gets the messages $f_{1l}(X_1^n)$ and $f_{2l}(X_2^n)$, and applies decoding function $\varphi_l^n(f_{1l}(X_1^n), f_{2l}(X_2^n))$ to obtain an estimate of the source S^n , denoted by $\hat{S}_l^n$, satisfying squared error distortion constraint D_l. The central receiver gets all the four descriptions and applies the function $\varphi_0^n(f_{11}(X_1^n), f_{21}(X_2^n), f_{12}(X_1^n), f_{22}(X_2^n))$ to get $\hat{S}_0^n$, satisfying the distortion constraint D_0. Throughout the chapter, we assume that $0 < D_0 < \min\{D_1, D_2\}$ and $\max\{D_1, D_2\} < \sigma^2_S$.

Definition 2.1. *We say that the tuple* $(R_{11}, R_{12}, R_{21}, R_{22}, D_1, D_2, D_0)$ *is* achievable *if there exist encoding functions* $f_{11}(X_1^n)$, $f_{12}(X_1^n)$, $f_{21}(X_2^n)$ *and* $f_{22}(X_2^n)$ *and decoding functions* $\varphi_1^n(f_{1l}(X_1^n), f_{2l}(X_2^n))$, $\varphi_2^n(f_{12}(X_1^n), f_{22}(X_2^n))$ *and* $\varphi_0^n(f_{11}(X_1^n), f_{21}(X_2^n), f_{12}(X_1^n), f_{22}(X_2^n))$ *such that*

$$R_{kl} \geq \lim_{n\to\infty} \frac{1}{n} \log |C_{kl}|, \quad k, l \in \{1, 2\}$$
$$D_l \geq \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(S_i - \hat{S}_{li})^2], \quad l \in \{0, 1, 2\}.$$

The vacationing-CEO problem generalizes the CEO problem with two sensors and the MD problem with two descriptions. In effect, the vacationing-CEO problem without the observation noise at one encoder and infinite noise variance at the other encoder reduces to the two description problem. If we omit one of the two stages by relaxing the corresponding distortion constraint, it reduces to a CEO problem. We briefly describe the main results of both these problems in the following.

2.2.2 The Quadratic Gaussian CEO Problem

Let $D_0 \geq \sigma_S^2$ and $D_2 \geq \sigma_S^2$. Therefore $R_{12} = R_{22} = 0$ and the resulting problem is the quadratic Gaussian CEO problem with distortion constraint D_1. Note that by imposing a trivial distortion constraint on any two out of the three receivers, it is possible to obtain a corresponding quadratic Gaussian CEO problem. The rate region for the CEO problem mentioned here consists of the set of all pairs (R_{11}, R_{21}) such that $\forall A \subset \{1, 2\}$, and $(r_1, r_2) \in \mathcal{F}$,

$$\sum_{k\in A} R_{k1} \geq \sum_{k\in A} r_k - \frac{1}{2} \log D_1 - \frac{1}{2} \log \left(\frac{1}{\sigma_S^2} + \sum_{k\in A^c} \frac{1 - e^{-2r_k}}{\sigma_{N_k}^2} \right),$$

where,

$$\mathcal{F} = \left\{ (r_1, r_2) : \frac{1}{D_1} = \frac{1}{\sigma_S^2} + \frac{1 - e^{-2r_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2r_2}}{\sigma_{N_2}^2} \right\}.$$

The optimal rate region is achieved by a Gaussian quantize and bin strategy at each encoder based on the Berger-Tung coding scheme. The proof of converse involves the idea of parametrizing the rate of quantization of the noise in the observations at each encoder. While the result as stated above, represents the rate region through parameters involving the rate of quantization of the noise, [87] obtains a sum rate characterization by parameterizing the distortion in the noise quantization. Neither of these techniques yield an obvious lower bound for the vacationing CEO problem. The lower bounding techniques for the CEO problem need to be combined in a non-trivial fashion with those for the MD problem to be able to obtain lower bounds for the vacationing CEO problem.

2.2.3 Gaussian Multiple Description

Let $\sigma_{N_1}^2 = 0$ and $\sigma_{N_2}^2 = \infty$. Therefore $R_{21} = R_{22} = 0$. As a result, we obtain the Gaussian two description problem with distortion constraints D_0, D_1 and D_2. Note that we can also obtain a similar Gaussian two description problem by setting $\sigma_{N_1}^2 = \infty$ and $\sigma_{N_2}^2 = 0$. The rate region of the former

Gaussian two description is given by the set of all pairs (R_{11}, R_{12}) such that

$$\begin{aligned} R_{11} &\geq \frac{1}{2}\log\frac{\sigma_S^2}{D_1} \\ R_{12} &\geq \frac{1}{2}\log\frac{\sigma_S^2}{D_2} \\ R_{11}+R_{12} &\geq \sup_{\sigma_Z^2\geq 0}\frac{1}{2}\log\frac{\sigma_S^2(\sigma_S^2+\sigma_Z^2)(D_0+\sigma_Z^2)}{D_0(D_1+\sigma_Z^2)(D_2+\sigma_Z^2)}. \end{aligned}$$

The Gaussian El Gamal - Cover coding strategy is optimal for this problem, while the converse involves the technique of supplying a noisy version of the source given which the two descriptions are independent. The extension of this converse technique to the vacationing-CEO problem is non trivial since it is not clear a priori whether a noisy version of the underlying source needs to be introduced or noisy versions of the encoder observations need to be introduced to induce conditional independence. This dilemma leads to two lower bounds on the sum rate of the vacationing-CEO problem, each of which is optimal in different distortion regimes.

2.2.4 Main Results

We now state the main results of this chapter of the dissertation. A brief description of the main theorems of the chapter is contained in Table 2.1. We first state two lower bounds on the sum rate and then present the sum rate achievable through the following theorems. The first lower bound is optimal in the low distortion regime (and achieved by a Gaussian scheme), while the second lower bound is optimal in the high distortion regime (and achieved by a time sharing scheme). We note that the time sharing strategy that is optimal

Table 2.1: Main theorems of the chapter

Theorem	Description
1	Lower bound on sum rate, optimal in the low distortion regime
2	Lower bound on sum rate, optimal in the high distortion regime
3	Gaussian achievable scheme, optimal in the low distortion regime
4	Gaussian achievable scheme with time sharing, optimal in the high distortion regime
5	Sum rate optimality using Theorem 1 and 3, in the low distortion regime
6	Sum rate optimality using Theorem 2 and 4, in the high distortion regime

in the 'high' distortion regime, involves a time sharing of Gaussian schemes, where each of the Gaussian schemes by themselves could potentially violate the distortion constraints at one of the individual receivers. In effect, these schemes are time shared in a manner such that all the distortion constraints are satisfied. Each lower bound involves an optimization problem over code parameters as in the proof of the CEO problem [45, 87]. However, we require a combination of parametrizations used in different proofs of the CEO problem for one of our lower bounds.

We now define a set $\mathcal{F}$, which contains all code parameters that satisfy the distortion constraints at all the receivers. For $k = 1, 2$, define the sets

$$\mathcal{F}_k = \{(d_1, d_2, t) : d_1, d_2, t \in [0, \infty) \\ \sigma^2_{N_k} e^{-2t} \leq \min\{d_1, d_2\} \quad \max\{d_1, d_2\} \leq \sigma^2_{N_k}\},$$

and

$$\mathcal{F} = \{(d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2) : (d_{k1}, d_{k2}, t_k) \in \mathcal{F}_k, k = 1, 2$$

$$\frac{1}{D_1} \leq \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d_{11}}{\sigma_{N_1}^4} - \frac{d_{21}}{\sigma_{N_2}^4} \tag{2.1}$$

$$\frac{1}{D_2} \leq \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d_{12}}{\sigma_{N_1}^4} - \frac{d_{22}}{\sigma_{N_2}^4} \tag{2.2}$$

$$\frac{1}{D_0} \leq \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2}\}. \tag{2.3}$$

We further define two sets, the first one in which all the distortion constraints are met with equality and the second one in which the central distortion is not met with equality. The objective of the definitions below is to consider an appropriate set for the lower bound optimization.

Let

$$\mathcal{P}_1 = \{(d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2) : (d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2) \in \mathcal{F}$$

$$\frac{1}{D_1} = \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d_{11}}{\sigma_{N_1}^4} - \frac{d_{21}}{\sigma_{N_2}^4} \tag{2.4}$$

$$\frac{1}{D_2} = \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d_{12}}{\sigma_{N_1}^4} - \frac{d_{22}}{\sigma_{N_2}^4} \tag{2.5}$$

$$\frac{1}{D_0} = \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2}\} \tag{2.6}$$

$$\mathcal{P}_2 = \{(d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2) : (d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2) \in \mathcal{F}$$

$$\frac{1}{D_1} = \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d_{11}}{\sigma_{N_1}^4} - \frac{d_{21}}{\sigma_{N_2}^4}$$

$$\frac{1}{D_2} = \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d_{12}}{\sigma_{N_1}^4} - \frac{d_{22}}{\sigma_{N_2}^4}$$

$$\frac{1}{D_0} < \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2}$$

$$\sigma_{N_1}^2 e^{-2t_1} = \min\{d_{11}, d_{12}\} \quad \sigma_{N_2}^2 e^{-2t_2} = \min\{d_{21}, d_{22}\}\}.$$

We denote

$$\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2.$$

Note that the definition of $\mathcal{P}$ imposes the restriction on the parameters to satisfy the individual distortion constraints with equality. The central distortion constraint may be satisfied with equality or the parameters satisfy $\sigma^2_{N_k} e^{-2t_k} = \min\{d_{k1}, d_{k2}\}$ for $k = 1, 2$. We also observe that $\mathcal{P} \subset \mathcal{F}$.

Let $k = 1, 2$ and $\sigma_Z^2 \geq 0$. Define, for $(d_1, d_2, t) \in \mathcal{F}_k$,

$$r_k(d_1, d_2, t, \sigma_Z^2) = t + \frac{1}{2}\log\frac{(\sigma^2_{N_k} + \sigma_Z^2)}{(d_1 + \sigma_Z^2)(d_2 + \sigma_Z^2)} + \frac{1}{2}\log(\sigma^2_{N_k} e^{-2t} + \sigma_Z^2). \quad (2.7)$$

We endow this quantity with its own notation for convenience and for being able to write the lower bound concisely, not because of its intrinsic importance. We now state the first lower bound.

Theorem 2.1. *The sum rate of the vacationing-CEO problem is lower bounded by*

$$R_{lb1} = \inf_{\tilde{p}\in\mathcal{P}} \sup_{\sigma_{Z_1}, \sigma_{Z_2} \in \mathbb{R}} r_1(d_{11}, d_{12}, t_1, \sigma^2_{Z_1}) + r_2(d_{21}, d_{22}, t_2, \sigma^2_{Z_2}) + \frac{1}{2}\log\frac{\sigma_S^4}{D_1 D_2}. \quad (2.8)$$

Proof. Please see Section 2.4.1. □

For the second lower bound on the sum rate, let

$$\mathcal{T} = \left\{(t_1, t_2) : t_1, t_2 \geq 0 \text{ and } \frac{1}{D_0} = \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma^2_{N_1}} + \frac{1 - e^{-2t_2}}{\sigma^2_{N_2}}\right\}.$$

Theorem 2.2. *The sum rate of the vacationing-CEO problem is lower bounded by*

$$R_{lb2} = \inf_{(t_1,t_2)\in\mathcal{T}} (t_1 + t_2) + \sup_{\sigma_Z\in\mathbb{R}} \frac{1}{2} \log \frac{\sigma_S^2(\sigma_S^2+\sigma_Z^2)(D_0+\sigma_Z^2)}{D_0(D_1+\sigma_Z^2)(D_2+\sigma_Z^2)}. \tag{2.9}$$

Proof. Please see Section 2.4.2. □

The next theorem concerns a Gaussian achievable scheme and is proved in Section 2.3. Here we only state the theorem. Let

$$\begin{aligned}\mathcal{U} = \{&(U_{11},U_{12},U_{21},U_{22}) : U_{kl} = X_k + W_{kl} \text{ for } k,l \in \{1,2\}, W_{kl} \sim \mathcal{N}(0,\sigma_{W_{kl}}^2),\\ &(U_{11},U_{12}) - X_1 - X_2 - (U_{21},U_{22}),\\ &\mathbb{E}[(S - \mathbb{E}[S|U_{1l},U_{2l}])^2] \le D_l \text{ for } l \in \{1,2\} \text{ and}\\ &\mathbb{E}[(S - \mathbb{E}[S|U_{11},U_{12},U_{21},U_{22}])^2] \le D_0\}.\end{aligned}$$

Note that W_{k1} and W_{k2} can be correlated for $k \in \{1,2\}$ in the above definition.

Theorem 2.3. *The sum rate achievable by a Gaussian scheme is given by*

$$\inf_{(U_{11},U_{12},U_{21},U_{22})\in\mathcal{U}} I(X_1,X_2;U_{11},U_{12},U_{21},U_{22}) + I(U_{11},U_{21};U_{12},U_{22}). \tag{2.10}$$

Proof. Please see Section 2.3 and Appendix A.1. □

Next, we state the sum rate achievable by time sharing the Gaussian scheme described above. Let

$$\begin{aligned}\mathcal{F}_U = \{&f(u_{11},u_{12},u_{21},u_{22}) : U_{kl} = X_k + W_{kl} \text{ for } k,l \in \{1,2\}, W_{kl} \sim \mathcal{N}(0,\sigma_{W_{kl}}^2),\\ &(U_{11},U_{12}) - X_1 - X_2 - (U_{21},U_{22})\}.\end{aligned}$$

$\mathcal{F}_U$ contains the set of all distributions on the auxiliaries such that they are Gaussian distributed while satisfying a Markov condition. Let T be a time sharing random variable in $\{1, 2, \ldots, \tau\}$.

Theorem 2.4. *The sum rate achievable by a Gaussian scheme with time sharing is given by*

$$\begin{aligned} \inf_{f(t)f(x_1,x_2)f(u_{11},u_{12},u_{21},u_{22}|t,x_1,x_2)} & I(X_1, X_2; U_{11}, U_{12}, U_{21}, U_{22}|T) \\ & + I(U_{11}, U_{21}; U_{12}, U_{22}|T) \qquad (2.11) \\ \textit{subject to } & f(u_{11}, u_{12}, u_{21}, u_{22}|t) \in \mathcal{F}_U \\ & \mathbb{E}[(S - \mathbb{E}[S|U_{1l}, U_{2l}, T])^2] \leq D_l \textit{ for } l \in \{1, 2\} \\ & \mathbb{E}[(S - \mathbb{E}[S|U_{11}, U_{12}, U_{21}, U_{22}, T])^2] \leq D_0. \end{aligned}$$

Since a convex combination of the achievable rates is also achievable (by choosing the convex combination such that the distortion constraints are also satisfied), the rate indicated in Equation (2.11) is achievable. The proof of this follows along a standard procedure as in the proof of Theorem 15.3.4 in [15].

The following theorem characterizes the sum rate in the *low distortion* regime, where the Gaussian scheme is sum rate optimal. Low distortion refers to the individual distortion constraints and a precise definition of the low distortion regime is provided in the theorem.

Theorem 2.5. *The sum rate of the vacationing-CEO problem with distortion constraints* (D_1, D_2, D_0) *such that* $\frac{1}{D_1} + \frac{1}{D_2} - \max\{\frac{1}{\sigma^2_{N_1}}, \frac{1}{\sigma^2_{N_2}}\} - \frac{1}{\sigma^2_S} \geq \frac{1}{D_0}$ *is given by*

$$\inf_{(U_{11},U_{12},U_{21},U_{22})\in\mathcal{U}} I(X_1,X_2;U_{11},U_{12},U_{21},U_{22}) + I(U_{11},U_{21};U_{12},U_{22}).$$

Proof. Please see Section 2.5. □

The next theorem establishes the sum rate in the *high distortion* regime. Let $R_0(D_0)$ denote the optimal sum rate required to achieve a distortion D_0 at Receiver 0. Mathematically, $R_0(D_0)$ is the sum rate of the vacationing-CEO problem with $D_1 = D_2 = \sigma_S^2$. The theorem establishes necessary and sufficient conditions for the sum rate of the vacationing-CEO problem to be equal to the sum rate of the central CEO. In other words, we establish conditions under which the 'no excess' sum rate is achievable.

Theorem 2.6. *The sum rate of the vacationing-CEO problem with distortion constraints* (D_1, D_2, D_0) *is equal to* $R_0(D_0)$ *if and only if* $D_1 + D_2 \geq \sigma_S^2 + D_0$.

Proof. Please see Section 2.6. □

We next compare the two bounds and the achievable strategy through a numerical example. We consider a symmetric problem with symmetric noise variances σ_N^2 at the two encoders and symmetric distortion constraints D at the individual receivers.

In Fig. 2.2, we plot the sum rate as a function of D for all $D_0 \leq D \leq \sigma_S^2$. The plot is generated for $\sigma_S^2 = 10$, $\sigma_N^2 = 2$ and $D_0 = 2$. Observe that the

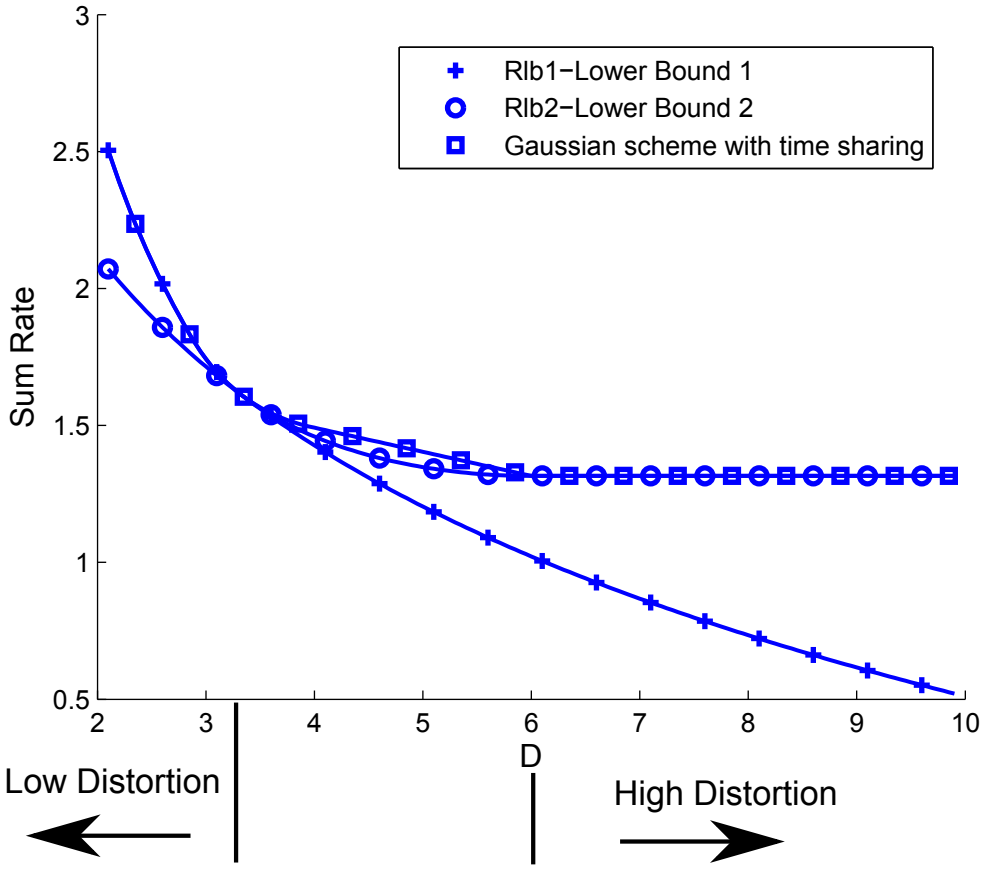


Figure 2.2: Sum rate of the symmetric vacationing-CEO problem

achievable sum rate matches with either of the lower bounds in the low distortion as well as the high distortion regimes. R_{lb1} is optimal for low distortions, corresponding to $D \leq 3.3$, while R_{lb2} is optimal for high distortions, corresponding to $D \geq 6$. While R_{lb1} is achievable by a Gaussian scheme, R_{lb2} is achievable by a simple time sharing scheme described in the proof of Theorem 2.6. We note that using the symmetry of the problem, one can show that the bounds match in the low distortion regime upto $D = 3.3$, which is confirmed numerically in the figure. This conclusion is a stronger version of Theorem 2.5 since the sufficient condition $D \leq 3.3$ contains the sufficient condition in Theorem 2.5.. For intermediate distortions, $3.3 \leq D \leq 6$, R_{lb2} is a better bound than R_{lb1}. We evaluate the sum rate achieved by the Gaussian scheme with time sharing based on Theorem 2.4. Note that the bounds determine the sum rate exactly over a wide range of D values, and elsewhere the bounds are very close. Further, the lower bound R_{lb1} is plotted by optimizing over $\sigma^2_{Z_1}$ and $\sigma^2_{Z_2}$ such that $\sigma^2_{Z_1} = \sigma^2_{Z_2}$. Also, the Gaussian scheme is evaluated by assuming symmetric forward test channel noise variances and symmetric correlations between these noises at the two encoders. Theorem 2.6 states that the 'no excess' sum rate $R_0(D_0)$ is not achievable for $D < 6$ and this can also be inferred from the figure.

Inspecting the two lower bounds, we see that R_{lb1} is tighter when the individual distortion constraints are more active and R_{lb2} is tighter when the central distortion constraint is more active. On the achievability side, in the low distortion regime, only a small degree of correlation between the quantiza-

tion that forms the two descriptions is required, and the Gaussian scheme can achieve this correlation despite the distributed nature of the encoding. Note that, in this regime, the sum rate does not always decouple into sum of the rates of the individual CEO problems. In the high distortion regime, a higher degree of correlation between the quantizations is required, which the Gaussian scheme without time sharing seems unable to achieve. The time sharing scheme is able to indirectly achieve the required correlation. In between, the difficulty seems to be determining whether there exist schemes that can realize the target correlation despite the distributed nature of the encoding. The difficulty in proving that the Gaussian schemes are optimal in the high distortion regime seems to stem from the restriction that the auxiliaries be chosen such that they simultaneously satisfy all the distortion constraints. Note that the time sharing scheme that is optimal, is a time sharing of Gaussian schemes where the Gaussian schemes themselves do not satisfy all the distortion constraints. These are time shared in a manner that satisfies all the distortion constraints and achieves the optimal sum rate.

In the following section, we discuss the Gaussian scheme whose achievable rates are characterized in Theorem 2.3.

2.3 A Gaussian scheme

We provide a brief sketch of the coding scheme below for discrete memoryless sources X_1^n and X_2^n with discrete auxiliary random variables $(U_{11}^n, U_{12}^n, U_{21}^n, U_{22}^n)$ satisfying the Markov condition $(U_{11}^n, U_{12}^n) - X_1^n - X_2^n -$

(U_{21}^n, U_{22}^n). We provide details of the Gaussian scheme later in this section.

2.3.1 Coding scheme for discrete memoryless sources

2.3.1.1 Codebook Generation

Encoder k, $k = 1, 2$, generates $2^{nR'_{k1}}$ U_{k1}^n and $2^{nR'_{k2}}$ U_{k2}^n such that U_{k1i} and U_{k2i} are generated i.i.d. according to the marginal of U_{k1} and U_{k2} respectively. $2^{nR'_{k1}}$ U_{k1}^n and $2^{nR'_{k2}}$ U_{k2}^n are binned into $2^{nR_{k1}}$ and $2^{nR_{k2}}$ bins respectively.

2.3.1.2 Encoding

Encoder k chooses the pair (U_{k1}^n, U_{k2}^n) jointly typical with X_k^n and transmits the respective bin indexes. There exists a pair (U_{k1}^n, U_{k2}^n) jointly typical with X_k^n with high probability if

$$\begin{aligned} R'_{k1} &> I(X_k; U_{k1}) \\ R'_{k2} &> I(X_k; U_{k2}) \\ R'_{k1} + R'_{k2} &> I(X_k; U_{k1}, U_{k2}) + I(U_{k1}; U_{k2}). \end{aligned} \tag{2.12}$$

This multiple description encoding scheme is similar to the scheme in [13]. Since $(U_{11}^n, U_{12}^n) - X_1^n - X_2^n - (U_{21}^n, U_{22}^n)$, by the Markov lemma (Lemma 14.8.1) in [15], we also have that $(U_{11}^n, U_{12}^n, U_{21}^n, U_{22}^n)$ are jointly typical.

2.3.1.3 Decoding at individual receivers

Receiver l, $l = 1, 2$, looks for U_{1l}^n and U_{2l}^n that are jointly typical in the bins corresponding to the bin indexes it receives. Receiver l will be able to

find unique codewords U_{1l}^n and U_{2l}^n that are jointly typical if

$$\begin{aligned} R_{1l} &> R'_{1l} - I(U_{1l}; U_{2l}) \\ R_{2l} &> R'_{2l} - I(U_{1l}; U_{2l}) \\ R_{1l} + R_{2l} &> R'_{1l} + R'_{2l} - I(U_{1l}; U_{2l}). \end{aligned} \tag{2.13}$$

The decoding scheme resembles the decoding in the Berger-Tung scheme [77].

2.3.1.4 Decoding at central receiver

Receiver 0 mimics the decoding at Receiver 1 and 2 to find jointly typical pairs (U_{11}^n, U_{21}^n) and (U_{12}^n, U_{22}^n) in the received bin indexes. Therefore, Receiver 0 will be able to find such unique codewords if the rates satisfy (2.13). Since $(U_{11}^n, U_{12}^n, U_{21}^n, U_{22}^n)$ are jointly typical.

Note that the equations in (2.12) and (2.13) represent the entire rate region achievable for the vacationing-CEO problem. In particular, the above scheme also achieves a sum rate equal to

$$I(X_1, X_2; U_{11}, U_{12}, U_{21}, U_{22}) + I(U_{11}, U_{21}; U_{12}, U_{22}),$$

the proof of which is supplied in Appendix A.1.

2.3.2 A Gaussian scheme

The coding scheme just described can be extended to Gaussian random variables using standard techniques [19, Chapter 3]. In the Gaussian scheme, the auxiliary random variables are chosen to be Gaussian and the estimates of

the desired source are constructed as follows. We define auxiliaries, U_{11},U_{12},U_{21} and U_{22} such that

$$U_{11} = X_1 + W_{11}$$
$$U_{12} = X_1 + W_{12}$$
$$U_{21} = X_2 + W_{21}$$
$$U_{22} = X_2 + W_{22},$$

where the vector $\mathbf{W} = (W_{11}, W_{12}, W_{21}, W_{22})$ is Gaussian distributed with mean zero and covariance matrix

$$\mathbf{K}_w = \begin{bmatrix} \sigma^2_{W_{11}} & -a_1 & 0 & 0 \\ -a_1 & \sigma^2_{W_{12}} & 0 & 0 \\ 0 & 0 & \sigma^2_{W_{21}} & -a_2 \\ 0 & 0 & -a_2 & \sigma^2_{W_{22}} \end{bmatrix}. \tag{2.14}$$

$\mathbf{K}_w$ is appropriately chosen to meet the distortion constraints at Receivers 1 and 2 and the central receiver. In effect $\mathbf{K}_w$ is chosen such that

$$\mathbb{E}\left[(S - \mathbb{E}\left[S|U_{1l}, U_{2l}\right])^2\right] \leq D_l, l \in \{1, 2\}$$
$$\mathbb{E}\left[(S - \mathbb{E}\left[S|U_{11}, U_{12}, U_{21}, U_{22}\right])^2\right] \leq D_0.$$

Receiver l generates an estimate of S^n, by constructing the minimum mean squared estimate (MMSE) $\mathbb{E}[S^n|U^n_{1l}, U^n_{2l}]$ for $l \in \{1, 2\}$. Receiver 0 constructs the MMSE estimate of S^n given by $\mathbb{E}[S^n|U^n_{11}, U^n_{12}, U^n_{21}, U^n_{22}]$. The auxiliaries chosen in this manner ensure that $(U_{11}, U_{12}, U_{21}, U_{22}) \in \mathcal{U}$.

We discuss the proofs of the lower bounds on the sum rate in the next section.

2.4 Lower Bounds on Sum Rate

In this section we present the proofs of lower bounds on the sum rate. We discuss the proof of Theorem 2.1 followed by Theorem 2.2

2.4.1 Proof of Theorem 2.1

The main idea of the proof is to parametrize the code in terms of the quantization of the noise and formulate the lower bound as an optimization problem over the parameters. On account of the multiple descriptions aspect of the problem, we also need to introduce additional random variables which represent noisy versions of the encoder observations. As a result, the lower bound also involves an optimization over the additional noise variance in the encoder observations. The proof presented here is different from [51] or [87] in that both parametrizations are required and different from [85] in that two noisy encoder observations are introduced as opposed to a single noisy observation of the underlying source in [85].

We begin by stating a lemma, which characterizes the parameters $\bar{p} = (d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2)$. Let C_{kl} denote the message from Encoder k to Receiver l for $k = 1, 2$ and $l = 1, 2$. Define

$$\begin{aligned} d_{11} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_{1i}|C_{11}, S^n) & d_{21} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_{2i}|C_{21}, S^n) \\ d_{12} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_{1i}|C_{12}, S^n) & d_{22} &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_{2i}|C_{22}, S^n) \\ t_1 &= \frac{1}{n} I(X_1^n; C_{11}, C_{12}|S^n) & t_2 &= \frac{1}{n} I(X_2^n; C_{21}, C_{22}|S^n). \end{aligned} \tag{2.15}$$

These parameters provide two different ways of characterizing the quantization of the noise at the encoders. While the conditional variance parameters represent the distortion in the noise due to quantization, the conditional mutual information parameter represents the rate of quantization of noise. We reiterate that we require both parametrizations in order to be able to obtain a lower bound and either one of them alone is not sufficient.

Lemma 2.7. *The parameters defined in (2.15) satisfy*

$$\bar{p} = (d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2) \in \mathcal{F}.$$

Proof. The proof that

$$\frac{1}{D_0} \leq \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2}$$

follows directly from Lemma 3.1 in [51]. Also, in Theorem 1 in [87], it is shown that

$$\frac{1}{D_l} \leq \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d_{1l}}{\sigma_{N_1}^4} - \frac{d_{2l}}{\sigma_{N_2}^4}$$

for $l = 1, 2$. By definition,

$$\begin{aligned} nt_k = I(X_k^n, C_{k1}, C_{k2}|S^n) &= h(X_k^n|S^n) - h(X_k^n|C_{k1}, C_{k2}, S^n) \\ &\geq \frac{n}{2} \log \sigma_{N_k}^2 - h(X_k^n|C_{kl}, S^n), l = 1, 2 \\ &\geq \frac{n}{2} \log \sigma_{N_k}^2 - \frac{n}{2} \log d_{kl}, l = 1, 2. \end{aligned}$$

Therefore for $k = 1, 2$,

$$\sigma_{N_k}^2 e^{-2t_k} \leq \min\{d_{k1}, d_{k2}\}.$$

Also, since $\mathbb{E}[N_k^n|C_{kl}]$ achieves a smaller mean squared error in N_k^n than any other estimator,

$$d_{kl} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(X_{ki}|C_{kl}, S^n) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}(N_{ki}|C_{kl}) \leq \sigma_{N_k}^2$$

for $k = 1, 2$ and $l = 1, 2$. This concludes the proof of the lemma. □

Let $\bar{p} \in \mathcal{F}$. Then $\Delta\mathcal{F}_{\bar{p}}$ is defined as

$$\begin{aligned}\Delta\mathcal{F}_{\bar{p}} = \{&\Delta\bar{p} = (\Delta d_{11}, \Delta d_{12}, \Delta d_{21}, \Delta d_{22}, -\Delta t_1, -\Delta t_2) : \\ &\Delta d_{11}, \Delta d_{12}, \Delta d_{21}, \Delta d_{22}, \Delta t_1, \Delta t_2 \in [0, \infty) \text{ and} \\ &(d_{11} + \Delta d_{11}, d_{12} + \Delta d_{12}, d_{21} + \Delta d_{21}, d_{22} + \Delta d_{22}, t_1 - \Delta t_1, t_2 - \Delta t_2) \in \mathcal{P}\}.\end{aligned}$$

The following lemma states that given a code, the code parameters can be perturbed so that they are elements of $\mathcal{P}$. Ultimately, the lower bound is also stated as an optimization problem of $\bar{p}$ over $\mathcal{P}$.

Lemma 2.8. $\Delta\mathcal{F}_{\bar{p}} \neq \phi \quad \forall \bar{p} \in \mathcal{F}$.

Proof. The lemma is proved as follows. Consider $\bar{p} \in \mathcal{F}$. Then we increase d_{11} and d_{12} by Δd_{11} and Δd_{12} until we meet the distortion constraints at individual receivers with equality or $d_{1l} + \Delta d_{1l} = \sigma_{N_1}^2$, $l = 1, 2$. In the former case, we satisfy the individual distortion constraints with equality. In the latter case, we now increase d_{21} and d_{22} by Δd_{21} and Δd_{22} until we meet the individual distortion constraints with equality. We will be able to find such Δd_{21} and Δd_{22} satisfying $d_{2l} + \Delta d_{2l} \leq \sigma_{N_2}^2$, $l = 1, 2$, since $D_l < \sigma_S^2$ for $l = 1, 2$. Now, we decrease t_1 by Δt_1 until the central distortion constraint

is met with equality or $\sigma_{N_1}^2 e^{-2(t_1-\Delta t_1)} = \min\{d_{11}+\Delta d_{11}, d_{12}+\Delta d_{12}\}$. In the former case, we satisfy the central distortion with equality. In the latter case, we decrease t_2 by Δt_2 until the central distortion constraint is met with equality or $\sigma_{N_2}^2 e^{-2(t_2-\Delta t_2)} = \min\{d_{21}+\Delta d_{21}, d_{22}+\Delta d_{22}\}$. Therefore $\forall \bar{p} \in \mathcal{F}$, $\Delta\mathcal{F}_{\bar{p}} \neq \phi$. □

We now turn to the proof of Theorem 2.1. By procedural steps, we have

$$
\begin{aligned}
& n(R_{11}+R_{21}+R_{12}+R_{22}) \\
& \quad \geq H(C_{11},C_{21}) + H(C_{12},C_{22}) \\
& \quad \geq H(C_{11},C_{21}) + H(C_{12},C_{22}) - H(C_{11},C_{21},C_{12},C_{22}) \\
& \qquad + H(C_{11},C_{21},C_{12},C_{22}) - H(C_{11},C_{21},C_{12},C_{22}|X_1^n,X_2^n) \\
& \quad = I(X_1^n,X_2^n;C_{11},C_{21},C_{12},C_{22}) + I(C_{11},C_{21};C_{12},C_{22}) \\
& \quad \overset{(a)}{=} I(S^n;C_{11},C_{21},C_{12},C_{22}) + I(X_1^n,X_2^n;C_{11},C_{21},C_{12},C_{22}|S^n) \\
& \qquad + I(C_{11},C_{21};C_{12},C_{22}) \\
& \quad \overset{(b)}{=} I(S^n;C_{11},C_{21},C_{12},C_{22}) + I(X_1^n;C_{11},C_{12}|S^n) + I(X_2^n;C_{21},C_{22}|S^n) \\
& \qquad + I(C_{11},C_{21};C_{12},C_{22}), \qquad (2.16)
\end{aligned}
$$

where (a) is true since $S^n - (X_1^n, X_2^n) - (C_{11}, C_{12}, C_{21}, C_{22})$ and (b) is true since $(C_{11}, C_{12}) - X_1^n - S^n - X_2^n - (C_{21}, C_{22})$.

Let $Y_{1i} = X_{1i} + Z_{1i}$ and $Y_{2i} = X_{2i} + Z_{2i}$, where Z_{1i} and Z_{2i} are i.i.d. Gaussians with mean zero and variance $\sigma_{Z_1}^2$ and $\sigma_{Z_2}^2$ for $i \in \{1, 2, \ldots, n\}$. Also, Z_{1i} and

Z_{2i} are independent of S_i, X_{1i} and X_{2i}. Now,

$$\begin{aligned}
&I(C_{11},C_{21};C_{12},C_{22}) \\
&\quad =H(C_{11},C_{21})+H(C_{12},C_{22})-H(C_{11},C_{21},C_{12},C_{22}) \\
&\quad =H(C_{11},C_{21})+H(C_{12},C_{22})-H(C_{11},C_{21},C_{12},C_{22}) \\
&\qquad -H(C_{11},C_{21}|S^n,Y_1^n,Y_2^n)-H(C_{12},C_{22}|S^n,Y_1^n,Y_2^n) \\
&\qquad +H(C_{11},C_{21},C_{12},C_{22}|S^n,Y_1^n,Y_2^n)+I(C_{11},C_{21};C_{12},C_{22}|Y_1^n,Y_2^n,S^n) \\
&\quad =I(S^n,Y_1^n,Y_2^n;C_{11},C_{21})+I(S^n,Y_1^n,Y_2^n;C_{12},C_{22}) \\
&\qquad -I(S^n,Y_1^n,Y_2^n;C_{11},C_{12},C_{21},C_{22})+I(C_{11},C_{21};C_{12},C_{22}|Y_1^n,Y_2^n,S^n) \\
&\quad \geq I(S^n,Y_1^n,Y_2^n;C_{11},C_{21})+I(S^n,Y_1^n,Y_2^n;C_{12},C_{22}) \\
&\qquad -I(S^n,Y_1^n,Y_2^n;C_{11},C_{12},C_{21},C_{22}). \qquad (2.17)
\end{aligned}$$

For $l=1,2$,

$$I(S^n,Y_1^n,Y_2^n;C_{1l},C_{2l})=I(S^n;C_{1l},C_{2l})+I(Y_1^n;C_{1l}|S^n)+I(Y_2^n;C_{2l}|S^n)$$

since $(Y_1^n,C_{1l})-S^n-(Y_2^n,C_{2l})$. By the definition of the rate distortion function for Gaussian random variables, $I(S^n;C_{1l},C_{2l})\geq\frac{n}{2}\log\frac{\sigma_S^2}{D_l}$ and $I(Y_k^n;C_{kl}|S^n)\geq\frac{n}{2}\log\frac{\sigma_{N_k}^2+\sigma_{Z_k}^2}{d_{kl}+\sigma_{Z_k}^2}$ for $k=1,2$. Therefore,

$$I(S^n,Y_1^n,Y_2^n;C_{1l},C_{2l})\geq\frac{n}{2}\log\frac{\sigma_S^2(\sigma_{N_1}^2+\sigma_{Z_1}^2)(\sigma_{N_2}^2+\sigma_{Z_2}^2)}{D_l(d_{1l}+\sigma_{Z_1}^2)(d_{2l}+\sigma_{Z_2}^2)}. \qquad (2.18)$$

Observe that

$$\begin{aligned}
&I(S^n, Y_1^n, Y_2^n; C_{11}, C_{12}, C_{21}, C_{22}) \\
&\quad = I(S^n; C_{11}, C_{21}, C_{12}, C_{22}) + I(Y_1^n, Y_2^n; C_{11}, C_{21}, C_{12}, C_{22} | S^n) \\
&\quad = I(S^n; C_{11}, C_{21}, C_{12}, C_{22}) + I(Y_1^n; C_{11}, C_{12} | S^n) \\
&\qquad + I(Y_2^n; C_{21}, C_{22} | S^n), \qquad (2.19)
\end{aligned}$$

where in the last step we used $(Y_1^n, C_{11}, C_{12}) - S^n - (Y_2^n, C_{21}, C_{22})$. Further, for $k = 1, 2$

$$\begin{aligned}
I(Y_k^n; C_{k1}, C_{k2} | S^n) &= -h(Y_k^n | S^n, C_{k1}, C_{k2}) + h(Y_k^n | S^n) \\
&\overset{(c)}{\leq} -\frac{n}{2} \log(e^{\frac{2}{n} h(X_k^n | S^n, C_{k1}, C_{k2})} + e^{\frac{2}{n} h(Z_k^n)}) + h(Y_k^n | S^n) \\
&= -\frac{n}{2} \log(e^{\frac{2}{n}(h(X_k^n | S^n) - I(X_k^n; C_{k1}, C_{k2} | S^n))} + e^{\frac{2}{n} h(Z_k^n)}) \\
&\quad + h(Y_k^n | S^n) \\
&= -\frac{n}{2} \log(\sigma_{N_k}^2 e^{-2t_k} + \sigma_{Z_k}^2) + \frac{n}{2} \log(\sigma_{N_k}^2 + \sigma_{Z_k}^2), \qquad (2.20)
\end{aligned}$$

where (c) follows from entropy power inequality (EPI). From (2.17), (2.18), (2.19) and (2.20),

$$\begin{aligned}
&I(C_{11}, C_{21}; C_{12}, C_{22}) \\
&\quad \geq \frac{n}{2} \log \frac{(\sigma_{N_1}^2 + \sigma_{Z_1}^2)(\sigma_{N_2}^2 + \sigma_{Z_2}^2)\sigma_S^4}{(d_{12} + \sigma_{Z_1}^2)(d_{22} + \sigma_{Z_2}^2)(d_{11} + \sigma_{Z_1}^2)(d_{21} + \sigma_{Z_2}^2) D_1 D_2} \\
&\qquad + \frac{n}{2} \log(\sigma_{N_1}^2 e^{-2t_1} + \sigma_{Z_1}^2)(\sigma_{N_2}^2 e^{-2t_2} + \sigma_{Z_2}^2) - I(S^n; C_{11}, C_{21}, C_{12}, C_{22})
\end{aligned}$$

Substituting the above in (2.16), we get

$$\begin{aligned}
&R_{11}+R_{21}+R_{12}+R_{22}\\
&\quad\geq t_1+t_2+\frac{1}{2}\log\frac{(\sigma_{N_1}^2+\sigma_{Z_1}^2)(\sigma_{N_2}^2+\sigma_{Z_2}^2)\sigma_S^4}{(d_{12}+\sigma_{Z_1}^2)(d_{22}+\sigma_{Z_2}^2)(d_{11}+\sigma_{Z_1}^2)(d_{21}+\sigma_{Z_2}^2)D_1D_2}\\
&\qquad+\frac{1}{2}\log(\sigma_{N_1}^2e^{-2t_1}+\sigma_{Z_1}^2)(\sigma_{N_2}^2e^{-2t_2}+\sigma_{Z_2}^2)\\
&\quad=r_1(d_{11},d_{12},t_1,\sigma_{Z_1}^2)+r_2(d_{21},d_{22},t_2,\sigma_{Z_2}^2)+\frac{1}{2}\log\frac{\sigma_S^4}{D_1D_2},
\end{aligned}\tag{2.21}$$

where the last equality is due to the definition in (2.7). From Lemma 2.7, we have $\bar{p}\in\mathcal{F}$. By Lemma 2.8, $\Delta\mathcal{F}_{\bar{p}}\neq\phi$. Let $\Delta\bar{p}\in\Delta\mathcal{F}_{\bar{p}}$. Note that $r_k(d_{k1},d_{k2},t_k,\sigma_{Z_k}^2)$ is decreasing in d_{k1} and d_{k2} and increasing in t_k for $k=1,2$. This implies that

$$r_k(d_{k1},d_{k2},t_k,\sigma_{Z_k}^2)\geq r_k(d_{k1}+\Delta d_{k1},d_{k2}+\Delta d_{k2},t_k-\Delta t_k,\sigma_{Z_k}^2)\quad\forall\bar{p}\in\mathcal{F}.$$

Therefore,

$$\begin{aligned}
R_{11}+R_{21}+R_{12}+R_{22}\geq&r_1(d_{11}+\Delta d_{11},d_{12}+\Delta d_{12},t_1-\Delta t_1,\sigma_{Z_1}^2)\\
&+r_2(d_{21}+\Delta d_{21},d_{22}+\Delta d_{22},t_2-\Delta t_2,\sigma_{Z_2}^2)\\
&+\frac{1}{2}\log\frac{\sigma_S^4}{D_1D_2}.
\end{aligned}$$

Note that by definition, $\bar{p}+\Delta\bar{p}\in\mathcal{P}$. Therefore,

$$\begin{aligned}
R_{11}+R_{21}+R_{12}+R_{22}\geq&\inf_{\bar{p}\in\mathcal{P}}\sup_{\sigma_{Z_1},\sigma_{Z_2}\in\mathbb{R}}r_1(d_{11},d_{12},t_1,\sigma_{Z_1}^2)+r_2(d_{21},d_{22},t_2,\sigma_{Z_2}^2)\\
&+\frac{1}{2}\log\frac{\sigma_S^4}{D_1D_2}.
\end{aligned}$$

2.4.2 Proof of Theorem 2.2

The proof of Theorem 2.2 is along the lines of [85] after the first few steps which proceed along the lines of [51]. The ideas in this proof are also similar to the lower bounding technique in [11]. Note that here we only introduce a single noisy version of the underlying Gaussian source as opposed to the previous theorem in which we introduced noisy versions of the two encoder observations to be able to obtain a lower bound on the sum rate. This enables the lower bound to be obtained such that schemes which do not meet the individual distortion constraints with equality could achieve this sum rate bound. The lower bound we obtain in the following is optimal in the high distortion regime.

Following the first few steps of the proof of Theorem 2.1, we get

$$\begin{aligned} n(R_{11}&+R_{21}+R_{12}+R_{22}) \\ &\geq I(S^n;C_{11},C_{21},C_{12},C_{22})+I(X_1^n;C_{11},C_{12}|S^n)+I(X_2^n;C_{21},C_{22}|S^n) \\ &\quad + I(C_{11},C_{21};C_{12},C_{22}) \\ &\geq nt_1+nt_2+I(S^n;C_{11},C_{21},C_{12},C_{22})+I(C_{11},C_{21};C_{12},C_{22}). \end{aligned} \tag{2.22}$$

and $(C_{11},C_{12},C_{21},C_{22})$ achieves D_0 in S, by introducing $Y^n=S^n+Z^n$, where $Z_i\sim\mathcal{N}(0,\sigma_Z^2)$, $i=1,\ldots,n$, i.i.d., we can lower bound the remaining terms in

(2.22) along the lines of proof of Lemma 2 of [85] as follows.

$$\begin{aligned}
&I(S^n; C_{11}, C_{21}, C_{12}, C_{22}) + I(C_{11}, C_{21}; C_{12}, C_{22}) \\
&\quad = I(S^n; C_{11}, C_{21}, C_{12}, C_{22}) + H(C_{11}, C_{21}) + H(C_{12}, C_{22}) \\
&\qquad - H(C_{11}, C_{12}, C_{21}, C_{22}) \\
&\quad \geq I(S^n; C_{11}, C_{21}, C_{12}, C_{22}) + H(C_{11}, C_{21}) + H(C_{12}, C_{22}) \\
&\qquad - H(C_{11}, C_{12}, C_{21}, C_{22}) - I(C_{11}, C_{21}; C_{12}, C_{22} | Y^n) \\
&\quad = h(S^n) + h(Y^n) - h(Y^n | C_{11}, C_{21}) - h(Y^n | C_{12}, C_{22}) \\
&\qquad + h(Y^n | C_{11}, C_{12}, C_{21}, C_{22}) - h(S^n | C_{11}, C_{12}, C_{21}, C_{22}) \qquad (2.23)
\end{aligned}$$

We have

$$h(S^n) = \frac{n}{2} \log(2\pi e)\sigma_S^2 \quad h(Y^n) = \frac{n}{2} \log(2\pi e)(\sigma_S^2 + \sigma_Z^2)$$

Now, since the code (C_{11}, C_{21}) achieves a distortion D_1, (C_{12}, C_{22}) achieves D_2, we have

$$h(Y^n | C_{11}, C_{21}) \leq \frac{n}{2} \log(2\pi e)(D_1 + \sigma_Z^2) \quad h(Y^n | C_{12}, C_{22}) \leq \frac{n}{2} \log(2\pi e)(D_2 + \sigma_Z^2).$$

Now,

$$
\begin{aligned}
&h(Y^n|C_{11},C_{21},C_{12},C_{22}) - h(S^n|C_{11},C_{21},C_{12},C_{22}) \\
&\qquad = h(Y^n|C_{11},C_{21},C_{12},C_{22}) - h(S^n|Z^n,C_{11},C_{21},C_{12},C_{22}) \\
&\qquad = h(Y^n|C_{11},C_{21},C_{12},C_{22}) - h(Y^n|Z^n,C_{11},C_{21},C_{12},C_{22}) \\
&\qquad = I(Y^n;Z^n|C_{11},C_{21},C_{12},C_{22}) \\
&\qquad = h(Z^n|C_{11},C_{21},C_{12},C_{22}) - h(Z^n|Y^n,C_{11},C_{21},C_{12},C_{22}) \\
&\qquad = h(Z^n) - h(Z^n|Y^n - \hat{S}_0^n,C_{11},C_{21},C_{12},C_{22}) \\
&\qquad \geq h(Z^n) - h(Z^n|Y^n - S_0^n) \\
&\qquad \geq \sum_{i=1}^{n} h(Z_i) - h(Z_i|Y_i - \hat{S}_{0i}) \\
&\qquad = \sum_{i=1}^{n} I(Z_i;Z_i + S_i - \hat{S}_{0i}) \\
&\qquad \geq \frac{n}{2}\log\frac{D_0+\sigma_Z^2}{D_0},
\end{aligned}
$$

where we use the property that Gaussian noise minimizes the mutual information [18]. Thus we get

$$
R_{11}+R_{12}+R_{21}+R_{22} \geq \inf_{(t_1,t_2)\in\mathcal{T}}(t_1+t_2) + \sup_{\sigma_Z\in\mathbb{R}}\frac{1}{2}\log\frac{\sigma_S^2(\sigma_S^2+\sigma_Z^2)(D_0+\sigma_Z^2)}{D_0(D_1+\sigma_Z^2)(D_2+\sigma_Z^2)}. \tag{2.24}
$$

In the following section, we show that Lower Bound 1 on the sum rate described earlier is achieved by the Gaussian scheme in the low distortion regime.

2.5 Proof of Theorem 2.5

While the main ideas of the proof of Theorem 2.5 are based on [85], there are certain differences. The different cases considered in the proof of [85] are in terms of the system parameters (distortion constraints and source variance), whereas here, the different cases considered are in terms of the parameters of the code. Further, we are only able to identify appropriate parameters $\sigma^2_{Z_1}$ and $\sigma^2_{Z_2}$ that enable the lower bound match with the achievable sum rate under the conditions stated in the theorem as opposed to being able to find such parameters that enable the lower bound match with the sum rate for all distortion constraints in [85].

Before we compare sum rate of the achievable scheme with the lower bound, we need two lemmas about parameters introduced earlier which will be used in the comparison. We will use the notation $\bar{p} = (d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2)$.

Lemma 2.9. *If* $\frac{1}{D_1} + \frac{1}{D_2} - \max\{\frac{1}{\sigma^2_{N_1}}, \frac{1}{\sigma^2_{N_2}}\} - \frac{1}{\sigma^2_S} \geq \frac{1}{D_0}$ *and* $\bar{p} \in \mathcal{P}$, *then*

$$d_{k1} + d_{k2} - \sigma^2_{N_k} e^{-2t_k} - \sigma^2_{N_k} \leq 0$$

for $k = 1, 2$.

Proof. Let $\bar{p} \in \mathcal{P}_1$. Since

$$\frac{1}{D_1} + \frac{1}{D_2} - \max\{\frac{1}{\sigma^2_{N_1}}, \frac{1}{\sigma^2_{N_2}}\} - \frac{1}{\sigma^2_S} \geq \frac{1}{D_0},$$

substituting for $\frac{1}{D_1}$, $\frac{1}{D_2}$ and $\frac{1}{D_0}$, from (2.4), (2.5) and (2.6) respectively, we get

$$\frac{d_{11} + d_{12} - \sigma^2_{N_1} - \sigma^2_{N_1} e^{-2t_1}}{\sigma^4_{N_1}} + \frac{d_{21} + d_{22} - \sigma^2_{N_2} - \sigma^2_{N_2} e^{-2t_2}}{\sigma^4_{N_2}} + \max\{\frac{1}{\sigma^2_{N_1}}, \frac{1}{\sigma^2_{N_2}}\} \leq 0.$$

Therefore at least one of $d_{11} + d_{12} - \sigma_{N_1}^2 - \sigma_{N_1}^2 e^{-2t_1} \leq 0$ or $d_{21} + d_{22} - \sigma_{N_2}^2 - \sigma_{N_2}^2 e^{-2t_2} \leq 0$. Let $d_{11} + d_{12} - \sigma_{N_1}^2 - \sigma_{N_1}^2 e^{-2t_1} \leq 0$. But since

$$\frac{d_{11} + d_{12} - \sigma_{N_1}^2 - \sigma_{N_1}^2 e^{-2t_1}}{\sigma_{N_1}^4} + \frac{d_{21} + d_{22} - \sigma_{N_2}^2 - \sigma_{N_2}^2 e^{-2t_2}}{\sigma_{N_2}^4} + \frac{1}{\sigma_{N_1}^2} \leq 0$$

$$\Rightarrow \frac{d_{11} + d_{12} - \sigma_{N_1}^2 e^{-2t_1}}{\sigma_{N_1}^4} + \frac{d_{21} + d_{22} - \sigma_{N_2}^2 - \sigma_{N_2}^2 e^{-2t_2}}{\sigma_{N_2}^4} \leq 0,$$

and $\sigma_{N_1}^2 e^{-2t_1} \leq \min\{d_{11}, d_{12}\}$, it follows that $d_{21} + d_{22} - \sigma_{N_2}^2 - \sigma_{N_2}^2 e^{-2t_2} \leq 0$. Similarly, we can start with $d_{21} + d_{22} - \sigma_{N_2}^2 - \sigma_{N_2}^2 e^{-2t_2} \leq 0$, and use

$$\frac{d_{11} + d_{12} - \sigma_{N_1}^2 - \sigma_{N_1}^2 e^{-2t_1}}{\sigma_{N_1}^4} + \frac{d_{21} + d_{22} - \sigma_{N_2}^2 - \sigma_{N_2}^2 e^{-2t_2}}{\sigma_{N_2}^4} + \frac{1}{\sigma_{N_2}^2} \leq 0,$$

to show that $d_{11} + d_{12} - \sigma_{N_1}^2 - \sigma_{N_1}^2 e^{-2t_1} \leq 0$. Therefore we have now shown that if $\bar{p} \in \mathcal{P}_1$, then $d_{k1} + d_{k2} - \sigma_{N_k}^2 e^{-2t_k} - \sigma_{N_k}^2 \leq 0$ for $k = 1, 2$.

Now, let $\bar{p} \in \mathcal{P}_2$. Therefore, $\sigma_{N_k}^2 e^{-2t_k} = \min\{d_{k1}, d_{k2}\}$, $k = 1, 2$. Since $\max\{d_{k1}, d_{k2}\} \leq \sigma_{N_k}^2$, it follows that

$$\begin{aligned} d_{k1} + d_{k2} - \sigma_{N_k}^2 - \sigma_{N_k}^2 e^{-2t_k} &= \min\{d_{k1}, d_{k2}\} + \max\{d_{k1}, d_{k2}\} - \sigma_{N_k}^2 - \sigma_{N_k}^2 e^{-2t_k} \\ &= \max\{d_{k1}, d_{k2}\} - \sigma_{N_k}^2 \\ &\leq 0. \end{aligned}$$

Thus for all $\bar{p} \in \mathcal{P}$, $d_{k1} + d_{k2} - \sigma_{N_k}^2 - \sigma_{N_k}^2 e^{-2t_k} \leq 0$, $k = 1, 2$. □

We make some more definitions which will help us show optimality of the Gaussian scheme in the low distortion regime. Let $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_k$ for $k = 1, 2$. Define

$$\alpha_{k0} = \frac{\sigma_{N_k}^4 e^{-2t_k}}{\sigma_{N_k}^2 - \sigma_{N_k}^2 e^{-2t_k}} \quad \alpha_{k1} = \frac{\sigma_{N_k}^2 d_{k1}}{\sigma_{N_k}^2 - d_{k1}} \quad \alpha_{k2} = \frac{\sigma_{N_k}^2 d_{k2}}{\sigma_{N_k}^2 - d_{k2}} \tag{2.25}$$

and

$$g_k(\beta) = \frac{1}{\alpha_{k0}+\beta} - \frac{1}{\alpha_{k1}+\beta} - \frac{1}{\alpha_{k2}+\beta}.$$

The solution to $g_k(\beta) = 0$, indicates the correlation in the noises of the Gaussian test channel to be chosen while generating the auxiliaries at each encoder. We also use this function to partition the space of parameters $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_k$. We require the partitioning to identify when and how the Gaussian scheme can meet the lower bound. Define,

$$\begin{aligned}
\mathcal{F}_{k1} =& \{(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_k : g_k(0) > 0 \text{ and } g_k(\sigma^2_{N_k}) \leq 0\} \\
\mathcal{F}_{k2} =& \{(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_k : g_k(0) \leq 0\} \\
\mathcal{F}_{k3} =& \{(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_k : g_k(\sigma^2_{N_k}) > 0\}.
\end{aligned}$$

Lemma 2.10. *For $k = 1, 2$,*

$$\mathcal{F}_k = \mathcal{F}_{k1} \cup \mathcal{F}_{k2} \cup \mathcal{F}_{k3}.$$

Moreover, if $\frac{1}{D_1} + \frac{1}{D_2} - \max\{\frac{1}{\sigma^2_{N_1}}, \frac{1}{\sigma^2_{N_2}}\} - \frac{1}{\sigma^2_S} \geq \frac{1}{D_0}$, then

$$\bar{p} \in \mathcal{P} \Rightarrow (d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k1} \cup \mathcal{F}_{k2}, k = 1, 2.$$

Proof. For every $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_k$, one of either $g_k(0) > 0$ and $g_k(\sigma^2_{N_k}) \leq 0$ or $g_k(0) \leq 0$ or $g_k(\sigma^2_{N_k}) > 0$ is true and therefore $\mathcal{F}_k = \mathcal{F}_{k1} \cup \mathcal{F}_{k2} \cup \mathcal{F}_{k3}$. From Lemma 2.9, $\bar{p} \in \mathcal{P}$ implies $d_{k1} + d_{k2} - \sigma^2_{N_k} e^{-2t_k} - \sigma^2_{N_k} \leq 0$ for $k = 1, 2$. However, $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k3}$ implies $g_k(\sigma^2_{N_k}) > 0$. This means that $d_{k1} + d_{k2} - \sigma^2_{N_k} e^{-2t_k} - \sigma^2_{N_k} > 0$. Therefore, $\bar{p} \in \mathcal{P}$ implies, $(d_{k1}, d_{k2}, t_k) \notin \mathcal{F}_{k3}$. Therefore,

$$\bar{p} \in \mathcal{P} \Rightarrow (d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k1} \cup \mathcal{F}_{k2}, k = 1, 2.$$

□

In order to show that the Gaussian scheme described in Section 2.3 achieves the lower bound on the sum rate, we parametrize the achievable sum rate now in the same manner as in the lower bound. Our objective is to write the achievable sum rate through similar parameters and prove that it is optimal. Define,

$$d'_{k1} = \mathrm{Var}(X_k|U_{k1}, S) = \frac{\sigma^2_{N_k}\sigma^2_{W_{k1}}}{\sigma^2_{N_k} + \sigma^2_{W_{k1}}} \tag{2.26}$$

$$d'_{k2} = \mathrm{Var}(X_k|U_{k2}, S) = \frac{\sigma^2_{N_k}\sigma^2_{W_{k2}}}{\sigma^2_{N_k} + \sigma^2_{W_{k2}}} \tag{2.27}$$

$$t'_k = I(X_k; U_{k1}, U_{k2}|S) = \frac{1}{2}\log\frac{\sigma^2_{N_k}(\sigma^2_{W_{k1}} + \sigma^2_{W_{k2}} + 2a_k) + \sigma^2_{W_{k1}}\sigma^2_{W_{k2}} - a_k^2}{\sigma^2_{W_{k1}}\sigma^2_{W_{k2}} - a_k^2}. \tag{2.28}$$

We can rewrite the last equation above as

$$\frac{1}{\frac{\sigma^2_{N_k}e^{-2t'_k}}{1-e^{-2t'_k}} + a_k} = \frac{1}{\sigma^2_{W_{k1}} + a_k} + \frac{1}{\sigma^2_{W_{k2}} + a_k}. \tag{2.29}$$

Let $\bar{p}' = (d'_{11}, d'_{12}, d'_{21}, d'_{22}, t'_1, t'_2)$ denote the parameters achieved by the Gaussian scheme. By definition of $(U_{11}, U_{12}, U_{21}, U_{22}) \in \mathcal{U}$, $\bar{p}' \in \mathcal{F}$. This means that the achievable parameters correspond to a Gaussian scheme that satisfies the distortion constraints (2.1), (2.2) and (2.3). We use the definition of functions in (2.7) and the parameters introduced above in the following lemma, relating them to the sum rate achievable by the Gaussian scheme.

Lemma 2.11. *For all* $(U_{11}, U_{12}, U_{21}, U_{22}) \in \mathcal{U}$ *and* $\sigma_{Z_k}^2 \geq 0$, $k \in \{1, 2\}$,

$$\begin{aligned} I(X_1, X_2; U_{11}, U_{12}, U_{21}, U_{22}) + I(U_{11}, U_{21}; U_{12}, U_{22}) \\ = r_1(d'_{11}, d'_{12}, t'_1, \sigma_{Z_1}^2) + r_2(d'_{21}, d'_{22}, t'_2, \sigma_{Z_2}^2) + \frac{1}{2}\log\frac{\sigma_S^4}{\delta_1\delta_2} \\ + I(U_{11}; U_{12}|S, Y_1) + I(U_{21}; U_{22}|S, Y_2), \end{aligned} \tag{2.30}$$

where

$$\begin{aligned} \frac{1}{\delta_1} &= \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d'_{11}}{\sigma_{N_1}^4} - \frac{d'_{21}}{\sigma_{N_2}^4} \\ \frac{1}{\delta_2} &= \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d'_{12}}{\sigma_{N_1}^4} - \frac{d'_{22}}{\sigma_{N_2}^4}, \end{aligned}$$

$Y_1 = X_1 + Z_1$ *and* $Y_2 = X_2 + Z_2$, Z_1 *and* Z_2 *are independent of both* X_1 *and* X_2 *and Gaussian distributed with mean zero and variance* $\sigma_{Z_1}^2$ *and* $\sigma_{Z_2}^2$ *respectively.*

This lemma is proved in Appendix A.2. We now show that the Gaussian scheme achieves the lower bound on the sum rate corresponding to every point $\bar{p} \in \mathcal{P}$ through the following lemma.

Lemma 2.12. *If* $\frac{1}{D_1} + \frac{1}{D_2} - \max\{\frac{1}{\sigma_{N_1}^2}, \frac{1}{\sigma_{N_2}^2}\} - \frac{1}{\sigma_S^2} \geq \frac{1}{D_0}$, *then for every* $\bar{p} \in \mathcal{P}$, *there exists an achievable* $\bar{p}' \in \mathcal{F}$ *and* $\sigma_{Z_k}^2 \geq 0$, $k = 1, 2$, *such that the sum rate achievable by the Gaussian scheme is*

$$r_1(d_{11}, d_{12}, t_1, \sigma_{Z_1}^2) + r_2(d_{21}, d_{22}, t_2, \sigma_{Z_2}^2) + \frac{1}{2}\log\frac{\sigma_S^4}{D_1 D_2}.$$

Proof. The proof closely follows the discussion in Section 5 in [85]. Let $\bar{p} = (d_{11}, d_{12}, d_{21}, d_{22}, t_1, t_2) \in \mathcal{P}$. Choosing $d'_{kl} = d_{kl}$ for $k = 1, 2$ and $l = 1, 2$, from

(2.4) and (2.5), we know that

$$\delta_1 = D_1 \quad \delta_2 = D_2. \tag{2.31}$$

From Lemma 2.10, we know that $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k1} \cup \mathcal{F}_{k2}$. By definition, $\mathcal{F}_{k1} \cap \mathcal{F}_{k2} = \phi$. We now consider two cases, $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k1}$ and $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k2}$.

2.5.1 Case 1: $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k1}$

Since $(d_{k1}, d_{k2}, t_k) \in \mathcal{P}_{k1}$, $g_k(0) > 0$ and $g_k(\sigma^2_{N_k}) \leq 0$. Therefore, there exists an $a_k^* \in (0, \sigma^2_{N_k}]$ that solves $g_k(a_k) = 0$. We set $a_k = a_k^*$. Recall that a_k is the correlation between W_{k1} and W_{k2} for $k \in \{1, 2\}$ as defined in (2.14). Further, $d'_{k1} = d_{k1}$ and $d'_{k2} = d_{k2}$ imply that $\sigma^2_{W_{k1}} = \alpha_{k1}$ and $\sigma^2_{W_{k2}} = \alpha_{k2}$. Therefore, we conclude from (2.29) and $g_k(a_k^*) = 0$ that $t'_k = t_k$. We now need to show that this choice of a_k^* is such that $\sigma^2_{W_{k1}} \sigma^2_{W_{k2}} \geq (a_k^*)^2$. Since $\alpha_{k0} \geq 0$ and $a_k^* \in (0, \sigma^2_{N_k}]$,

$$\alpha_{k0} + a_k^* \geq a_k^* \Rightarrow \frac{1}{\alpha_{k0} + a_k^*} \leq \frac{1}{a_k^*}.$$

Since $g(a_k^*) = 0$,

$$\begin{aligned}
&\frac{1}{\sigma^2_{W_{k1}} + a_k^*} + \frac{1}{\sigma^2_{W_{k2}} + a_k^*} \leq \frac{1}{a_k^*} \\
\Rightarrow &\frac{1}{\sigma^2_{W_{k1}} + a_k^*} \leq \frac{1}{a_k^*} - \frac{1}{\sigma^2_{W_{k2}} + a_k^*} \\
\Rightarrow &\frac{1}{\sigma^2_{W_{k1}} + a_k^*} \leq \frac{\sigma^2_{W_{k2}}}{a_k^*(\sigma^2_{W_{k2}} + a_k^*)} \\
\Rightarrow &\sigma^2_{W_{k1}} + a_k^* \geq \frac{(a_k^*)^2}{\sigma^2_{W_{k2}}} + a_k^* \\
\Rightarrow &\sigma^2_{W_{k1}} \sigma^2_{W_{k2}} \geq (a_k^*)^2.
\end{aligned}$$

Moreover, trivially,

$$r_k(d_{k1}, d_{k2}, t_k, \sigma^2_{Z_k}) = r_k(d'_{k1}, d'_{k2}, t'_k, \sigma^2_{Z_k})$$

Also,

$$\mathrm{Var}(U_{kl}|S, Y_k) = \sigma^2_{N_k} + \sigma^2_{W_{kl}} - \frac{\sigma^4_{N_k}}{\sigma^2_{N_k} + \sigma^2_{Z_k}}, l = 1, 2$$

$$\mathrm{Cov}(U_{k1}, U_{k2}|S, Y_k) = \sigma^2_{N_k} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2_{W_{k1}} & -a_k \\ -a_k & \sigma^2_{W_{k2}} \end{bmatrix} - \frac{\sigma^4_{N_k}}{\sigma^2_{N_k} + \sigma^2_{Z_k}} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The off diagonal entries in $\mathrm{Cov}(U_{k1}, U_{k2}|S, Y_k)$ are zero if

$$\sigma^2_{N_k} - a_k = \frac{\sigma^4_{N_k}}{\sigma^2_{N_k} + \sigma^2_{Z_k}}.$$

By choosing $\sigma^2_{Z_k} = \frac{a_k \sigma^2_{N_k}}{\sigma^2_{N_k} - a_k}$ in this case,

$$\mathrm{Var}(U_{k1}|S, Y_k)\,\mathrm{Var}(U_{k2}|S, Y_k) = |\,\mathrm{Cov}(U_{k1}, U_{k2}|S, Y_k)|$$

and $I(U_{k1}; U_{k2}|S, Y_k) = 0$. Note that we are allowed to choose $\sigma^2_{Z_k} = \frac{a_k \sigma^2_{N_k}}{\sigma^2_{N_k} - a_k}$ since $a_k \in (0, \sigma^2_{N_k}]$ in this case.

2.5.2 Case 2: $(d_{k1}, d_{k2}, t_k) \in \mathcal{F}_{k2}$

In this case, we set $a_k = 0$ in (2.28) and achieve the corresponding t'_k. Since, $d'_{k1} = d_{k1}$ and $d'_{k2} = d_{k2}$, we have $\sigma^2_{W_{k1}} = \alpha_{k1}$ and $\sigma^2_{W_{k2}} = \alpha_{k2}$. It follows from (2.25) and (2.29) that

$$\frac{1}{\sigma^2_{N_k}} + \frac{1}{\alpha_{k1}} + \frac{1}{\alpha_{k2}} = \frac{1}{\sigma^2_{N_k} e^{-2t'_k}}.$$

Since $g_k(0) \leq 0$, this implies that

$$\begin{aligned} \frac{1}{\alpha_{k0}} &\leq \frac{1}{\alpha_{k1}} + \frac{1}{\alpha_{k2}} \\ \Rightarrow \frac{1}{\alpha_{k0}} + \frac{1}{\sigma^2_{N_k}} &\leq \frac{1}{\sigma^2_{N_k}} + \frac{1}{\alpha_{k1}} + \frac{1}{\alpha_{k2}} \\ \Rightarrow \frac{1}{\sigma^2_{N_k} e^{-2t_k}} &\leq \frac{1}{\sigma^2_{N_k}} + \frac{1}{\alpha_{k1}} + \frac{1}{\alpha_{k2}} \\ &= \frac{1}{\sigma^2_{N_k} e^{-2t'_k}}. \end{aligned}$$

Therefore, we get that $t_k \leq t'_k$. By achieving t'_k instead of t_k, we still satisfy the central distortion constraint for the original problem and also ensure $(d'_{k1}, d'_{k2}, t'_k) \in \mathcal{F}_k$. Further, we choose $\sigma^2_{Z_k} = 0$ in this case. Therefore

$$r_k(d_{k1}, d_{k2}, t_k, 0) = \frac{1}{2} \log \frac{\sigma^4_{N_k}}{d_{k1} d_{k2}} = r_k(d_{k1}, d_{k2}, t'_k, 0),$$

Moreover, since $\sigma^2_{Z_k} = 0$ and $a_k = 0$

$$I(U_{k1}; U_{k2} | S, Y_k) = I(U_{k1}; U_{k2} | S, X_k) = 0.$$

The lemma follows from the cases considered above. □

Therefore, from Theorem 2.1 and Lemma 2.12, for every $\bar{p} \in \mathcal{P}$, there exists an achievable $\bar{p}' \in \mathcal{F}$ such that the sum rate achievable by the Gaussian scheme is equal to the lower bound on the sum rate. This proves the optimality of the Gaussian scheme for the sum rate of the vacationing-CEO problem in the 'low distortion' regime.

2.6 Proof of Theorem 2.6

We discuss both the achievability and converse arguments for the proof of Theorem 2.6 in the following. The achievability argument of the proof of Theorem 2.6 involves a time sharing strategy based on Theorem 2.4. Note that while a time sharing strategy is not necessary in [46], we are able to show the achievability of $R_0(D_0)$ for $D_1 + D_2 \geq \sigma_S^2 + D_0$ only by using the time sharing scheme. The converse proof involves introducing a random variable which is a noisy version of the underlying Gaussian source, similar to [46]. In order to prove that for the sum rate of the vacationing CEO problem to be equal to the sum rate of the central CEO, the distortion constraints should satisfy the given inequality, we assume the opposite inequality and prove a contradiction. We closely follow the proof in [46] and adapt it to the CEO problem setting in order to prove the contradiction. The definition of the parameters t_1 and t_2 in (2.15) continue to hold in this section as well.

We first show the 'if' part of the proof. The achievability of the sum rate in the high distortion regime is a time sharing argument. The set of all rates achievable by time sharing is stated in Theorem 2.4. Here we explicitly describe one such time sharing scheme to achieve the no excess rate. Let C_1 and C_2 be optimal block codes of length n from Encoder 1 and 2 respectively to the central receiver that achieve the sum rate for the central receiver (sum rate of the CEO problem) corresponding to parameters $\bar{p}$. We now let $C_k^{\alpha m}$ be the code C_k sent over αm blocks and $C_k^{(1-\alpha)m}$ be the code sent over $(1-\alpha)m$ blocks. Thus, Receiver 1 gets the optimal code from both encoders for α fraction of

the blocks while Receiver 2 gets the optimal code from both encoders for $1-\alpha$ fraction of the blocks. Without loss of generality, let us assume $D_1 \leq D_2$. Now, we choose α such that

$$D_1 = \alpha D_0 + (1-\alpha)\sigma_S^2.$$

Such a choice exists since $D_0 \leq D_1 \leq \sigma_S^2$. Receiver 1 reconstructs α fraction of blocks of source samples from the received code while the remaining $(1-\alpha)$ fraction of blocks is estimated as zero. On the other hand, Receiver 2 reconstructs $1-\alpha$ fraction of the block of source samples from the code it receives and estimates the remaining $(1-\alpha)$ fraction as zero. Clearly, the distortion achieved by Receiver 1 is $\alpha D_0 + (1-\alpha)\sigma_S^2 = D_1$. Receiver 2 achieves a distortion $(1-\alpha)D_0 + \alpha\sigma_S^2$. However, since $D_1 + D_2 \geq \sigma_S^2 + D_0$, the distortion achieved by Receiver 2 is less than D_2. The distortion at the central receiver is always satisfied. This time sharing scheme achieves a sum rate equal to the sum rate of the central CEO given by [45],

$$R_0(D_0) = \inf_{(t_1,t_2)\in\mathcal{T}} \frac{1}{2}\log\frac{\sigma_S^2}{D_0} + t_1 + t_2.$$

Observe that the discussion above concerns a time sharing of two schemes, each of which violates the distortion constraint at one of the individual receivers. However, the two schemes are time shared in a manner that results in a scheme that satisfies the distortion constraints at all the three receivers while achieving the sum rate of the central CEO.

While it is easy to see that the sum rate of the vacationing-CEO problem is lower bounded by the sum rate of the central CEO, $R_0(D_0)$, we also

recover this bound using Theorem 2.2. By Theorem 2.2, the sum rate is lower bounded by

$$\inf_{(t_1,t_2)\in\mathcal{T}}(t_1+t_2)+\sup_{\sigma_Z\in\mathbb{R}}\frac{1}{2}\log\frac{\sigma_S^2(\sigma_S^2+\sigma_Z^2)(D_0+\sigma_Z^2)}{D_0(D_1+\sigma_Z^2)(D_2+\sigma_Z^2)}.$$

Now, allowing $\sigma_Z^2\to\infty$, we get that the sum rate is lower bounded by

$$R_0(D_0)=\inf_{(t_1,t_2)\in\mathcal{T}}\frac{1}{2}\log\frac{\sigma_S^2}{D_0}+t_1+t_2,$$

thus proving the optimality of the time sharing scheme.

We now show the 'only if' part of the proof. The proof is largely based on Ozarow's proof technique [46]. Since $(R_{11},R_{12},R_{21},R_{22},D_1,D_2,D_0)$ is achievable, $\forall\alpha,\beta>0$, there exists $N(\alpha,\beta)$ such that $\forall n>N(\alpha,\beta)$, we have

$$\begin{aligned}
&n(R_{11}+R_{12}+R_{21}+R_{22}+\alpha)\\
&\quad\geq H(C_{11},C_{21})+H(C_{12},C_{22})\\
&\quad= H(C_{11},C_{21})+H(C_{12},C_{22})-H(C_{11},C_{12},C_{21},C_{22})\\
&\qquad+H(C_{11},C_{12},C_{21},C_{22})-H(C_{11},C_{12},C_{21},C_{22}|X_1^n,X_2^n,S^n)\\
&\quad= I(S^n,X_1^n,X_2^n;C_{11},C_{12},C_{21},C_{22})+I(C_{11},C_{21};C_{12},C_{22})\\
&\quad= I(S^n;C_{11},C_{12},C_{21},C_{22})+I(X_1^n;C_{11},C_{12}|S^n)\\
&\qquad+I(X_2^n;C_{21},C_{22}|S^n)+I(C_{11},C_{21};C_{12},C_{22}) \qquad (2.32)\\
&\quad= I(S^n;C_{11},C_{21})+I(S^n;C_{12},C_{22})+I(C_{11},C_{21};C_{12},C_{22}|S^n)\\
&\qquad+nt_1+nt_2 \qquad (2.33)\\
&\quad\geq \frac{n}{2}\log\frac{\sigma_S^2}{D_1+\beta}+\frac{n}{2}\log\frac{\sigma_S^2}{D_2+\beta}+nt_1+nt_2, \qquad (2.34)
\end{aligned}$$

where we obtain (2.33) by using the definitions in (2.15) and the equality

$$\begin{aligned}
&I(C_{11}, C_{21}; C_{21}, C_{22}) + I(S^n; C_{11}, C_{12}, C_{21}, C_{22}) \\
&\quad = H(C_{11}, C_{21}) + H(C_{12}, C_{22}) - H(C_{11}, C_{12}, C_{21}, C_{22}|S^n) \\
&\quad = I(S^n; C_{11}, C_{21}) + I(S^n; C_{12}, C_{22}) + H(C_{11}, C_{21}|S^n) + H(C_{12}, C_{22}|S^n) \\
&\qquad - H(C_{11}, C_{12}, C_{21}, C_{22}|S^n) \\
&\quad = I(S^n; C_{11}, C_{21}) + I(S^n; C_{12}, C_{22}) + I(C_{11}, C_{21}; C_{12}, C_{22}|S^n).
\end{aligned}$$

Let (t_1^*, t_2^*) be the optimal solution to the problem,

$$\begin{aligned}
&\min \quad t_1 + t_2 \\
&\text{subject to } \frac{1}{D_0} = \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2} \text{ and } t_1, t_2 \geq 0,
\end{aligned} \tag{2.35}$$

and $(\tilde{t}_1, \tilde{t}_2)$ be the optimal solution to the problem,

$$\begin{aligned}
&\min \quad t_1 + t_2 \\
&\text{subject to } \frac{1}{D_0 + \beta} = \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2} \text{ and } t_1, t_2 \geq 0, .
\end{aligned} \tag{2.36}$$

The optimization problems listed above in (2.35) and (2.36) may be equivalently written with inequality constraints as

$$\begin{aligned}
&\min \quad t_1 + t_2 \\
&\text{subject to } \frac{1}{D_0} \leq \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2} \text{ and } t_1, t_2 \geq 0,
\end{aligned}$$

and

$$\begin{aligned}
&\min \quad t_1 + t_2 \\
&\text{subject to } \frac{1}{D_0 + \beta} \leq \frac{1}{\sigma_S^2} + \frac{1 - e^{-2t_1}}{\sigma_{N_1}^2} + \frac{1 - e^{-2t_2}}{\sigma_{N_2}^2} \text{ and } t_1, t_2 \geq 0.
\end{aligned} \tag{2.37}$$

Therefore, we see that $\tilde{t}_1+\tilde{t}_2 \leq t_1^*+t_2^*$. Therefore, we define $\delta = t_1^*+t_2^*-\tilde{t}_1-\tilde{t}_2$ which satisfies, $\delta \geq 0$. Since $(R_{11}, R_{12}, R_{21}, R_{22}, D_1, D_2, D_0)$ achieves a sum rate $R_0(D_0)$,

$$R_{11} + R_{12} + R_{21} + R_{22} = \frac{1}{2}\log\frac{\sigma_S^2}{D_0} + t_1^* + t_2^*. \tag{2.38}$$

We define,

$$\begin{aligned}
\tilde{D}_1 &= D_1 + \beta \\
\tilde{D}_2 &= D_2 + \beta \\
\Pi &= (\sigma_S^2 - D_1)(\sigma_S^2 - D_2) \\
\Delta &= D_1 D_2 - \sigma_S^4 \exp\left[-2(R_{11} + R_{12} + R_{21} + R_{22} - t_1^* - t_2^*)\right] \\
\tilde{\Pi} &= (\sigma_S^2 - \tilde{D}_1)(\sigma_S^2 - \tilde{D}_2) \\
\tilde{\Delta} &= \tilde{D}_1 \tilde{D}_2 - \sigma_S^4 \exp\left[-2(R_{11} + R_{12} + R_{21} + R_{22} + \alpha - \tilde{t}_1 - \tilde{t}_2)\right].
\end{aligned}$$

In the following, we consider β, small enough such that $\tilde{D}_1 < \sigma_S^2$ and $\tilde{D}_2 < \sigma_S^2$. Therefore $\tilde{\Pi} > 0$. Note that, by (2.34), $\tilde{\Delta} \geq 0$. Also, note that $\tilde{t}_1 + \tilde{t}_2$ is continuous in β. This can be seen by observing that the optimization problem in (2.37) is convex and the KKT conditions imply that the unique optimizers are continuous in β. Therefore, we also have $\Delta \geq 0$.

Now, we prove the desired result by contradiction. Let us assume $\Pi > \Delta$.

Since, $\Pi > \Delta$ and

$$\begin{aligned}
\tilde{\Pi} =& \Pi + (D_1 + D_2 - 2\sigma_S^2)\beta + \beta^2 \\
\tilde{\Delta} =& \Delta + (D_1 + D_2)\beta + \beta^2 \\
&+ \sigma_S^4 \exp\left[-2(R_{11} + R_{12} + R_{21} + R_{22} - t_1^* - t_2^*)\right](1 - e^{-2(\alpha+\delta)}) \\
\Rightarrow \tilde{\Pi} - \tilde{\Delta} =& \Pi - \Delta - 2\sigma_S^2\beta \\
&- \sigma_S^4 \exp\left[-2(R_{11} + R_{12} + R_{21} + R_{22} - t_1^* - t_2^*)\right](1 - e^{-2(\alpha+\delta)}),
\end{aligned}$$

there exists $\alpha_0, \beta_0 > 0$, such that for all $\alpha < \alpha_0$, $\beta < \beta_0$, $\tilde{\Pi} > \tilde{\Delta}$. This is true due to the continuity of δ in β which follows from the continuity of $\tilde{t}_1 + \tilde{t}_2$ in β. In the subsequent discussion, we consider $\alpha < \alpha_0$ and $\beta < \beta_0$.

Now, from (2.32), and data processing inequality,

$$\begin{aligned}
I(S^n; \hat{S}_1^n, \hat{S}_2^n) &\leq n(R_{11} + R_{12} + R_{21} + R_{22} + \alpha) - n\tilde{t}_1 - n\tilde{t}_2 - I(\hat{S}_1^n; \hat{S}_2^n) \\
I(S^n; \hat{S}_0^n) &\leq n(R_{11} + R_{12} + R_{21} + R_{22} + \alpha) - n\tilde{t}_1 - n\tilde{t}_2 - I(\hat{S}_1^n; \hat{S}_2^n).
\end{aligned} \tag{2.39}$$

By definition of the distortion rate function for Gaussian sources and the above,

$$\begin{aligned}
D_0 + \beta &\geq \sigma_S^2 \exp\left(-\frac{2}{n} I(S^n; \hat{S}_0^n)\right) \\
&\geq \sigma_S^2 \exp\left(-2(R_{11} + R_{12} + R_{21} + R_{22} + \alpha - \tilde{t}_1 - \tilde{t}_2) + \frac{2}{n} I(\hat{S}_1^n; \hat{S}_2^n)\right).
\end{aligned} \tag{2.40}$$

Let $Y^n = S^n + Z^n$, where Z_i is i.i.d. Gaussian for $i = 1, \ldots, n$, with mean zero and variance σ_Z^2 and independent of S^n. Hence,

$$\begin{aligned}
I(\hat{S}_1^n; \hat{S}_2^n) &= I(Y^n; \hat{S}_1^n) + I(Y^n; \hat{S}_2^n) + I(\hat{S}_1^n; \hat{S}_2^n | Y^n) - I(Y^n; \hat{S}_1^n, \hat{S}_2^n) \\
&\geq I(Y^n; \hat{S}_1^n) + I(Y^n; \hat{S}_2^n) - I(Y^n; \hat{S}_1^n, \hat{S}_2^n).
\end{aligned} \tag{2.41}$$

Since $\hat{S}_l^n$ achieves a distortion $\tilde{D}_l$ for $l = 1, 2$, it follows by the point to point rate distortion function that

$$\frac{1}{n} I(Y^n; \hat{S}_l^n) \geq \frac{1}{2} \log \frac{\sigma_S^2 + \sigma_Z^2}{\tilde{D}_l + \sigma_Z^2}. \tag{2.42}$$

Now,

$$\begin{aligned} I(Y^n; \hat{S}_1^n, \hat{S}_2^n) &= h(Y^n) - h(Y^n | \hat{S}_1^n, \hat{S}_2^n) \\ &= \frac{n}{2} \log 2\pi e(\sigma_S^2 + \sigma_Z^2) - h(Y^n | \hat{S}_1^n, \hat{S}_2^n) \\ &\leq \frac{n}{2} \log 2\pi e(\sigma_S^2 + \sigma_Z^2) - \frac{n}{2} \log \left[\exp\left(\frac{2}{n} h(S^n | \hat{S}_1^n, \hat{S}_2^n) \right) + 2\pi e \sigma_Z^2 \right]. \end{aligned} \tag{2.43}$$

By equation (2.39),

$$h(S^n | \hat{S}_1^n, \hat{S}_2^n) \geq \frac{n}{2} \log 2\pi e \sigma_S^2 - n(R_{11} + R_{12} + R_{21} + R_{22} + \alpha - \tilde{t}_1 - \tilde{t}_2) + I(\hat{S}_1^n; \hat{S}_2^n).$$

Combining the above with (2.43),

$$\begin{aligned} & I(Y^n; \hat{S}_1^n, \hat{S}_2^n) \\ & \leq \frac{n}{2} \log 2\pi e(\sigma_S^2 + \sigma_Z^2) \\ & \quad - \frac{n}{2} \log 2\pi e \Bigg[\sigma_Z^2 \\ & \qquad + \sigma_S^2 \exp\left(-2(R_{11} + R_{12} + R_{21} + R_{22} + \alpha - \tilde{t}_1 - \tilde{t}_2) + \frac{2}{n} I(\hat{S}_1^n; \hat{S}_2^n) \right) \Bigg]. \end{aligned} \tag{2.44}$$

Letting $\gamma = \exp\left(\frac{2}{n} I(\hat{S}_1^n; \hat{S}_2^n) \right)$, from (2.41), (2.42) and (2.44), we get,

$$\gamma \geq \frac{(\sigma_S^2 + \sigma_Z^2)(\sigma_S^2 \exp\left[-2(R_{11} + R_{12} + R_{21} + R_{22} + \alpha - \tilde{t}_1 - \tilde{t}_2) \right] \gamma + \sigma_Z^2)}{(\sigma_Z^2 + \tilde{D}_1)(\sigma_Z^2 + \tilde{D}_2)}.$$

Alternatively,

$$\gamma \geq \frac{\sigma_Z^2(\sigma_S^2+\sigma_Z^2)}{\sigma_Z^4+\frac{\sigma_Z^2}{\sigma_S^2}(\sigma_S^4+\tilde{\Delta}-\tilde{\Pi})+\tilde{\Delta}}.$$

Maximizing the RHS with respect to σ_Z^2, we get

$$\sigma_Z^2 = \frac{\sqrt{\tilde{\Delta}}\sigma_S^2}{\sqrt{\tilde{\Pi}}-\sqrt{\tilde{\Delta}}}$$
$$\gamma \geq \frac{\sigma_S^4}{\sigma_S^4-(\sqrt{\tilde{\Pi}}-\sqrt{\tilde{\Delta}})^2}.$$

Note that the σ_Z^2 specified above is valid since $\tilde{\Pi} > \tilde{\Delta}$. Thus, from (2.40), we conclude that,

$$D_0+\beta$$
$$\geq \sigma_S^2 \exp\left[-2(R_{11}+R_{12}+R_{21}+R_{22}+\alpha+\delta-t_1^*-t_2^*)\right] \frac{\sigma_S^4}{\sigma_S^4-(\sqrt{\tilde{\Pi}}-\sqrt{\tilde{\Delta}})^2}.$$

Since the above inequality is true for all $\alpha < \alpha_0$ and $\beta < \beta_0$, by allowing $\alpha \to 0$ and $\beta \to 0$, we get

$$D_0 \geq \sigma_S^2 \exp\left[-2(R_{11}+R_{12}+R_{21}+R_{22}-t_1^*-t_2^*)\right] \frac{\sigma_S^4}{\sigma_S^4-(\sqrt{\Pi}-\sqrt{\Delta})^2}. \tag{2.45}$$

We again use the continuity of δ in β in the last step.

Further, we have $\sigma_S^4-(\sqrt{\Pi}-\sqrt{\Delta})^2 > 0$. This can be shown as follows:

$$\sigma_S^2(D_1+D_2)-2D_1D_2+\sigma_S^2D_0 > -2\sqrt{\Delta\Pi}$$
$$\Rightarrow \sigma_S^4 > \sigma_S^4-\sigma_S^2(D_1+D_2)+2D_1D_2-\sigma_S^2D_0-2\sqrt{\Delta\Pi}$$
$$\Rightarrow \sigma_S^4 > (\sqrt{\Pi}-\sqrt{\Delta})^2,$$

where we write $\Delta = D_1 D_2 - \sigma_S^2 D_0$ (when the sum rate is equal to $R_0(D_0)$) and the first step is true since $\sigma_S^2(D_1 + D_2) \geq 2D_1 D_2$. Therefore, from (2.45) and (2.38) we have,

$$(\sqrt{\Pi} - \sqrt{\Delta})^2 \leq 0,$$

which is a contradiction since $\Pi > \Delta$ and $\Pi, \Delta \geq 0$. Therefore the case $\Pi > \Delta$ is not possible under the condition of no excess rate. Thus, $\Pi \leq \Delta$.

If $\Pi \leq \Delta$, then we have

$$(\sigma_S^2 - D_1)(\sigma_S^2 - D_2) \leq D_1 D_2 - \sigma_S^4 \exp\left[-2(R_{11} + R_{12} + R_{21} + R_{22} - t_1^* - t_2^*)\right]$$

$$\Rightarrow \sigma_S^4 - \sigma_S^2(D_1 + D_2) + D_1 D_2 \leq D_1 D_2 - \sigma_S^2 D_0 \tag{2.46}$$

$$\Rightarrow D_1 + D_2 \geq \sigma_S^2 + D_0, \tag{2.47}$$

where (2.46) follows from (2.38). Hence, the 'only if' part of the theorem is also shown.

Chapter 3

Multiterminal Source Coding Through a Relay

3.1 Introduction

In recent years, there has been considerable progress made in improving our understanding of multi-terminal source coding problems, particularly for cases with Gaussian sources and squared distortion constraints. The rate region for the Gaussian CEO problem as introduced in [79] was characterized in [45, 51]. Further, the rate region for the Gaussian two-terminal compression problem was characterized in [83]. A common achievable strategy in this body of work is to use quantization (using a random coding argument) followed by Slepian-Wolf binning. This achievable strategy, although a powerful tool for many multi-terminal source coding problems, is not a one-size-fits-all solution. For distributed function computation as studied in [33], lattice based achievable strategies were found to play a key role in obtaining better achievable rates than random coding and binning arguments, especially for linear difference functions.

Indeed, lattices have played an important role in compression, both in theory [21, 100] and practice [58]. The single source single destination quadratic Gaussian rate distortion theorem has been re-established using lat-

tices [100]. Lattices provide us with structural properties that aid both compression and communication problems. Lattices have had a particularly meaningful impact on joint source-channel coding problems, in the form of computational codes [41], for multiple access [42, 71], broadcast channels [72], channels with state [32] and the exchange channel [96].

In this chapter, we investigate the problem of multi-terminal source coding through a relay. Depicted in Fig. 3.1, this setting comprises of two terminals with two distinct sources S_1 and S_2, each of which are noisy versions of an underlying Gaussian source S. It is desired that S be reconstructed within a distortion constraint at the decoder. Unlike a conventional multi-terminal problem, an intermediate relay node is present that can re-encode and forward information to the decoder. At first glance, this looks like a minor variation on the original multi-terminal problem, but we find that the presence of a relay makes a significant difference in terms of achievability strategies necessary for this setting. In particular, we find that:

- the structure of lattices proves particularly useful in achieving "good" rates for a multi-terminal source coding problem through a relay; and that,
- for classes of multi-terminal source coding problems through a relay, our upper and lower bounds on sum rate are within 1/2 bit of each other.

Thus, the sum rate of a class of multi-terminal source coding problems through a relay can be characterized to within 1/2 bit by using lattices with co-

operative outer bounds on the system. Note that relay-assisted source coding problems have received considerable attention recently, where a single source is being relayed from a source to a destination under different assumptions on the availability of side information [16, 47, 78]. Lattices have not proved essential to any of these existing results on relay-assisted point-to-point source coding problems as yet. It is the combination of a multi-terminal setting with relays where lattices are found useful in this chapter of the dissertation.

The rest of this chapter is organized as follows. The next section presents a mathematical framework for the multi-terminal source coding problem through a relay. Section 3.3 presents a lower bound on the sum rate for this problem while Section 3.4 presents an achievable strategy using lattices and compares the lower bound with the rates achieved.

3.2 System Model

First, a quick note on notation. We use capitals for random variables and the corresponding small case for their realizations. Z^n represents an n-length vector of random variables. We use boldface to denote matrices. All logarithms are with respect to base 2. $\sigma^2_{X|Y}$ denotes $\mathbb{E}\left[(X - \mathbb{E}[X|Y])^2\right]$.

The system model is depicted in Fig. 3.1. Consider two noisy versions of an underlying Gaussian source $\{S_i\}_{i=1}^n$, denoted by $\{S_{1i}\}_{i=1}^n$ and $\{S_{2i}\}_{i=1}^n$, where

$$S_{ki} = S_i + N_{ki}, k \in \{1, 2\}, i \in \{1, 2, \ldots, n\},$$

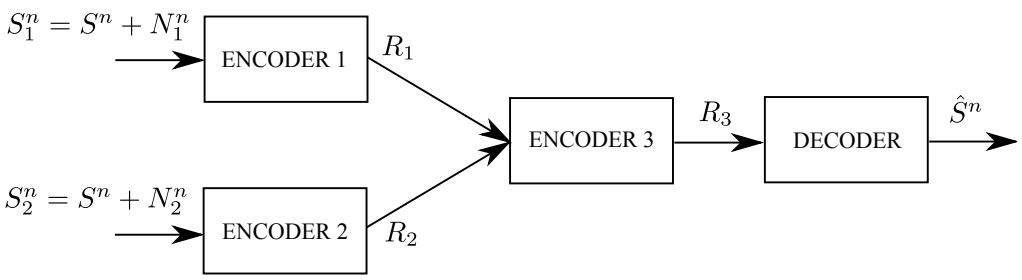


Figure 3.1: System model for multiterminal source coding through a relay

and $S_i \sim \mathcal{N}(0, \sigma_S^2)$, $N_{ki} \sim \mathcal{N}(0, \sigma_{N_k}^2)$, S_i, N_{1i} and N_{1i} are independent of each other and independent across the index i. Encoders 1 and 2 observe $\{S_{1i}\}_{i=1}^n$ and $\{S_{2i}\}_{i=1}^n$ respectively and transmit their compression indices to a relay (Encoder 3). The relay in turn sends an index to the decoder/CEO. Mathematically, Encoders 1 and 2 apply encoding functions given by

$$f_k : \mathbb{R}^n \to \{1, 2, \ldots, 2^{nR_k}\}, k \in \{1, 2\}$$

while the relay applies an encoding function given by

$$f_3 : \{1, 2, \ldots, 2^{nR_1}\} \times \{1, 2, \ldots, 2^{nR_2}\} \to \{1, 2, \ldots, 2^{nR_3}\}.$$

Encoder k, $k \in \{1, 2\}$, sends the index $C_k = f_k(S_k^n)$ to the relay. The relay sends the index $C_3 = f_3(C_1, C_2)$ to the CEO. The rate of transmission from Encoders 1, 2 and 3 are R_1, R_2 and R_3 respectively, by definition. The CEO applies a decoding function given by

$$g : \{1, 2, \ldots, 2^{nR_3}\} \to \mathbb{R}^n.$$

The goal is to recover the underlying source S within a squared error distortion

constraint D. Therefore $\hat{S}^n = g(C_3)$ needs to satisfy

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(S_i - \hat{S}_i)^2] \leq D.$$

Definition 3.1. *We say that the tuple (R_1, R_2, R_3, D) is achievable if there exist encoding functions f_1, f_2 and f_3 and decoding function g such that $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(S_i - \hat{S}_i)^2] \leq D$.*

The objective is to find the set all of achievable rate distortion tuples (R_1, R_2, R_3, D). Note that in the absence of the relay, i.e. when $R_3 = \infty$, the problem reduces to the well known CEO problem [79].

3.3 Lower Bound on Sum Rate

In order to lower bound the sum rate of the problem, we lower bound R_3 and the sum of R_1 and R_2. We state an outer bound on the rate region through the following theorem.

Theorem 3.1. *If (R_1, R_2, R_3, D) is* achievable*, then*

$$\begin{aligned} R_1 + R_2 \geq \quad & \min \quad \frac{1}{2}\log\frac{\sigma_S^2}{D} + r_1 + r_2 \\ & \textit{subject to } \frac{1}{\sigma_S^2} + \frac{1-2^{-2r_1}}{\sigma_{N_1}^2} + \frac{1-2^{-2r_1}}{\sigma_{N_2}^2} = \frac{1}{D}, \\ R_3 \geq \frac{1}{2}\log & \frac{\sigma_S^4(\sigma_{N_1}^2 + \sigma_{N_2}^2)}{D(\sigma_{N_1}^2\sigma_{N_2}^2 + \sigma_S^2\sigma_{N_1}^2 + \sigma_S^2\sigma_{N_2}^2) - \sigma_S^2\sigma_{N_1}^2\sigma_{N_2}^2}. \end{aligned}$$

Proof. We first bound the sum of R_1 and R_2 by allowing the relay and the CEO to cooperate. In effect, the pair (R_1, R_2) is bounded by the rate region

of the CEO problem with two sensors and distortion constraint D. Therefore the sum rate is lower bounded by

$$\begin{aligned} \min \quad & \frac{1}{2}\log\frac{\sigma_S^2}{D} + r_1 + r_2 \\ \text{subject to } & \frac{1}{\sigma_S^2} + \frac{1-2^{-2r_1}}{\sigma_{N_1}^2} + \frac{1-2^{-2r_1}}{\sigma_{N_2}^2} = \frac{1}{D}. \end{aligned}$$

Let

$$\Delta = \frac{1}{\frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2}}$$

and

$$\lambda = \frac{1}{\frac{1}{\Delta} - \frac{1}{D}}$$

Then the optimal r_1 and r_2 are given by

$$r_1^* = \frac{1}{2}\log^+\frac{2\lambda}{\sigma_{N_1}^2} \quad r_2^* = \frac{1}{2}\log^+\frac{2\lambda}{\sigma_{N_2}^2},$$

where $\log^+(x) = \max\{\log x, 0\}$ and the sum rate is lower bounded by [45],

$$R_1 + R_2 \geq \frac{1}{2}\log\frac{\sigma_S^2}{D} + r_1^* + r_2^* \tag{3.1}$$

Now, in order to bound R_3, we allow Encoders 1 and 2 to cooperate with the relay, Encoder 3. Therefore, the rate R_3 is bounded by the rate distortion function of the estimate of the source S given S_1 and S_2. Note that Δ equals the variance of the error in estimating S given S_1 and S_2. Therefore,

$$\begin{aligned} R_3 \geq & \frac{1}{2}\log\frac{\sigma_S^2 - \Delta}{D - \Delta} \\ = & \frac{1}{2}\log\frac{\sigma_S^4(\sigma_{N_1}^2 + \sigma_{N_2}^2)}{D(\sigma_{N_1}^2\sigma_{N_2}^2 + \sigma_S^2\sigma_{N_1}^2 + \sigma_S^2\sigma_{N_2}^2) - \sigma_S^2\sigma_{N_1}^2\sigma_{N_2}^2}. \end{aligned} \tag{3.2}$$

□

The bounds on the $R_1 + R_2$ and R_3 can be used to generate a lower bound on the sum rate of the system, $R_1 + R_2 + R_3$. In the next section, we discuss the achievability of different coding schemes that achieve close to the lower bound on the sum rate.

3.4 Lattice Coding based Achievability

On account of the well known tension between whether the relay should forward or compress [16], we describe two lattice coding schemes corresponding to the relay computing and forwarding or compressing and forwarding. We refer to the two strategies discussed below as *compute and forward* and *compress and forward*. The use of the lattices is crucial to achieving the rates due to *compute and forward*, while the *compress and forward* strategy can be used to achieve rates either by using lattices or random codes. Further, *compute and forward* is better than merely forwarding the codeword indices from the relay. The structure of lattices allows for efficient rates in performing the computation when compared to random codes. We briefly define terminologies related to lattice codes before presenting the lattice coding scheme.

A lattice of dimension n is defined as the set

$$\Lambda = \{x^n = z^n \mathbf{G} : z^n \in \mathbb{Z}^n\}$$

where $\mathbf{G} \in \mathbb{R}^{n \times n}$ is the generator matrix and $\mathbb{Z}$ is the set of all integers. The quantized value of $x^n \in \mathbb{R}^n$, $Q_\Lambda(x^n) = \text{argmin}_{r^n \in \Lambda} \|x^n - r^n\|_2$. The fundamental Voronoi region of Λ is defined as $\mathcal{V}_0 = \{x^n \in \mathbb{R}^n : Q_\Lambda(x^n) = 0\}$. The second

moment of a lattice is defined as

$$\sigma^2(\Lambda) = \frac{\int_{\mathcal{V}_0} \|x^n\|_2^2 dx}{n \int_{\mathcal{V}_0} dx}$$

while the normalized second moment is defined as

$$G(\Lambda) = \frac{\sigma^2(\Lambda)}{(\int_{\mathcal{V}_0} dx)^{\frac{2}{n}}}.$$

Also, let $x^n \bmod \Lambda = x^n - Q_\Lambda(x^n)$.

A lattice is said to be 'good' for channel coding if for an i.i.d sequence of Gaussian random variables $\{Z_i\}_{i=1}^n$ with mean zero and variance $\sigma^2(\Lambda)$,

$$\lim_{n\to\infty} \Pr(Z^n \in \mathcal{V}_0) = 1.$$

The lattice is 'good' for source coding if the normalized second moment satisfies the property,

$$\lim_{n\to\infty} G(\Lambda) = \frac{1}{2\pi e}.$$

The proof of existence of lattices that are simultaneously 'good' for both source as well as channel coding and their constructions can be found in [21]. We now present the lattice schemes. Let Λ_{11}, Λ_{12}, Λ_2 and Λ_3 be 'good' source and channel coding lattices of dimension n. Let the second moments of the lattices be $\sigma^2(\Lambda_{11})$, $\sigma^2(\Lambda_{12})$, $\sigma^2(\Lambda_2)$ and $\sigma^2(\Lambda_3)$. The lattice schemes proposed here are more along the lines of nested lattices used for multiterminal source coding in [101]. They differ from the way lattices have been used in [31, 33, 42] for other problems. We use Λ_3 for signal processing at the relay.

3.4.1 Compute and forward

In the following scheme, we let Encoder 1 and 2 operate at the optimal sum rate of the CEO problem. In order to achieve this, we allow Encoder 1 to transmit at a rate $I(S_1;U_1)$ and Encoder 2 to transmit at a rate $I(S_2;U_2|U_1)$, where $U_1 = S_1+W_1$ and $U_2 = S_2+W_2$ are distributed according to the optimal auxiliaries in the respective forward test channels [45]. The optimal choice of the variances of W_1 and W_2 satisfies

$$\sigma^2_{W_k} = \frac{\sigma^2_{N_k}}{\frac{2}{\sigma^2_{N_k}\left(\frac{1}{\Delta}-\frac{1}{D}\right)}-1}. \tag{3.3}$$

We assume that $\sigma^2_{W_k} > 0$ for $k \in \{1,2\}$ since, otherwise the problem reduces to a simple cascade problem [16]. While the achievability of the above rates using quantization followed by random binning is fairly trivial, we discuss the achievability using lattice codes which allows their use at the relay. Let the minimum mean squared error (MMSE) estimate of S given U_1 and U_2 be $\beta_1 U_1 + \beta_2 U_2$ and $S_3 = \beta_1 S_1 + \beta_2 S_2$. We choose the second moments of the lattices as:

$$\begin{aligned}
\sigma^2(\Lambda_1) &= \beta_1^2\sigma^2_{W_1}\\
\sigma^2(\Lambda_{21}) &= \beta_2^2\sigma^2_{W_2}\\
\sigma^2(\Lambda_{22}) &= \beta_2^2(\sigma^2_{W_2}+\sigma^2_{S_2|U_1})\\
\sigma^2(\Lambda_3) &= \sigma^2_{S_3}+\beta_1^2\sigma^2_{W_1}+\beta_2^2\sigma^2_{W_2}.
\end{aligned}$$

In the following, we assume that $\beta_1^2\sigma^2_{W_1} \le \beta_2^2\sigma^2_{W_2}$. If this is not satisfied, we achieve the rate pair $(I(S_1;U_1|U_2), I(S_2;U_2))$ by swapping the structure of the

coding schemes at Encoder 1 and 2, thereby achieving the same sum rate. The lattices are chosen with a nesting structure given by $\Lambda_3 \subset \Lambda_{22} \subset \Lambda_{21} \subseteq \Lambda_1$ when $\beta_1^2\sigma_{W_1}^2 \leq \beta_2^2\sigma_{W_2}^2$.

Encoder 1 sends

$$C_1 = Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n),$$

while Encoder 2 sends

$$C_2 = Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n) \bmod \Lambda_{22},$$

where Z_1^n and Z_2^n are independent dithers that are uniformly distributed in the fundamental Voronoi region of the lattices Λ_1 and Λ_{21} respectively. Further, the dithers are known at the relay and the receiver. Therefore, the rates achieved by the encoders are given by [100],

$$\begin{aligned} R_1 &= \frac{1}{2}\log\frac{\sigma_{S_1}^2 + \sigma_{W_1}^2}{\sigma_{W_1}^2} = I(S_1; U_1) \\ R_2 &= \frac{1}{2}\log\frac{\sigma_{W_2}^2 + \sigma_{S_2|U_1}^2}{\sigma_{W_2}^2} = I(S_2; U_2|U_1). \end{aligned}$$

Let the MMSE estimate of S_2 given U_1 be αU_1 and let $\alpha' = \frac{\beta_2}{\beta_1}\alpha$. Now the relay sends

$$\begin{aligned} C_3 = [&Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) + [[Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n)] \bmod \Lambda_{22} - Z_2^n \\ &-\alpha' Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) + \alpha' Z_1^n] \bmod \Lambda_{22} \\ &+\alpha' Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) - \alpha' Z_1^n + Z_2^n] \bmod \Lambda_3. \end{aligned} \tag{3.4}$$

Since $\sigma^2(\Lambda_{22}) = \beta_2^2(\sigma_{W_2}^2 + \sigma_{S_2|U_1}^2)$, with high probability, we have,

$$C_3 = [Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) + Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n)] \bmod \Lambda_3.$$

Note that the signal processing in (3.4) is required in order to extract the lattice point sent from Encoder 2 using the lattice point sent from Encoder 1. Further, we refer to the coding strategy thus employed as *compute and forward* since the codeword sent from the relay is just the sum of the codewords from Encoder 1 and 2 and no compression is performed at the relay. The modulo operation with respect to Λ_3 allows for the computation of rate achieved by the relay given by,

$$R_3 = \frac{1}{2}\log\frac{\sigma^2(\Lambda_3)}{\sigma^2(\Lambda_1)} = \frac{1}{2}\log\frac{\sigma^2_{S_3} + \beta_1^2\sigma^2_{W_1} + \beta_2^2\sigma^2_{W_2}}{\beta_1^2\sigma^2_{W_1}}. \tag{3.5}$$

Note that in the rate expression above, we consider the finer lattice among Λ_1 and Λ_{21} when lattice points belonging to the respective lattices are summed up. The decoder performs the following signal processing:

$$\begin{aligned} R^n &= [C_3 - Z_1^n - Z_2^n] \bmod \Lambda_3 \\ &= [\beta_1 S_1^n + \beta_2 S_2^n + V_1^n + V_2^n] \bmod \Lambda_3, \end{aligned}$$

where

$$\begin{aligned} V_1^n &= Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) - (\beta_1 S_1^n + Z_1^n) \\ V_2^n &= Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n) - (\beta_2 S_2^n + Z_2^n), \end{aligned} \tag{3.6}$$

with $\sigma^2_{V_1} = \beta_1^2\sigma^2_{W_1}$ and $\sigma^2_{V_2} = \beta_2^2\sigma^2_{W_2}$. Also, V_1^n and V_2^n are independent of each other and S_1^n and S_2^n [100]. Since $\sigma^2(\Lambda_3) = \sigma^2_{S_3} + \beta_1^2\sigma^2_{W_1} + \beta_2^2\sigma^2_{W_2}$, with high probability, we have

$$R^n = \beta_1 S_1^n + \beta_2 S_2^n + V_1^n + V_2^n.$$

The linear estimate of S^n given R^n is

$$\hat{S}^n = \frac{(\beta_1 + \beta_2)\sigma_S^2}{\sigma_{S_3}^2 + \beta_1^2\sigma_{W_1}^2 + \beta_2^2\sigma_{W_2}^2} R^n = R^n,$$

and results in a distortion

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(S_i - \hat{S}_i)^2] = D.$$

By substituting (3.3) in (3.5) we can calculate the resulting rate from the relay. Further, sufficient conditions on system parameters can be obtained such that the rate achieved is within a desired constant gap of the lower bound. In the following, we calculate the rate R_3 for the symmetric case, i.e. when $\sigma_{N_1}^2 = \sigma_{N_2}^2 = \sigma_N^2$, to get

$$R_3 = \frac{1}{2}\log\frac{4\sigma_S^4}{\sigma_N^2(D - \sigma_S^2) + 2\sigma_S^2 D}.$$

Comparing with (3.2), we see that the gap is bounded by 1/2 bit.

3.4.2 Compress and forward

We now discuss a coding strategy where the relay does perform compression. The objective of considering this strategy is to investigate whether it is possible to achieve a rate R_3 smaller than that achievable using the *compute and forward* strategy by allowing for rate R_1 and R_2 to be higher than the optimal rates for the CEO problem. Let U_1' and U_2' be the associated auxiliaries distributed according to $U_1' = S_1 + W_1'$ and $U_2' = S_2 + W_2'$, where $W_1' \sim \mathcal{N}(0, \sigma_{W_1'}^2)$ and $W_2' \sim \mathcal{N}(0, \sigma_{W_2'}^2)$. Recall that β_1 and β_2 are the optimal

linear coefficients corresponding to the MMSE of estimate of S given U_1 and U_2. Now, a sum rate of

$$R_1 + R_2 = I(S_1; U_1') + I(S_2; U_2'|U_1')$$

is achievable by letting

$$\begin{aligned}
\sigma^2(\Lambda_1) &= \beta_1^2 \sigma^2_{W_1'} \\
\sigma^2(\Lambda_{21}) &= \beta_2^2 \sigma^2_{W_2'} \\
\sigma^2(\Lambda_{22}) &= \beta_2^2 (\sigma^2_{W_2'} + \sigma^2_{S_2|U_1'}),
\end{aligned}$$

and

$$\begin{aligned}
C_1 &= Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) \\
C_2 &= Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n) \bmod \Lambda_{22},
\end{aligned}$$

By choosing $\sigma^2_{W_k'} \leq \sigma^2_{W_k}$, $k \in \{1, 2\}$, we achieve a higher $R_1 + R_2$ than the optimal sum rate of the CEO problem. Now, let

$$\sigma^2(\Lambda_3) = \sigma^2_{W_3}.$$

The relay performs similar processing to (3.4) without the modulo operation with respect to Λ_3 to get $Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) + Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n)$, where Z_1^n and Z_2^n are dithers defined as before. Let Z_3^n be an independent dither that is uniformly distributed in the fundamental Voronoi region of lattice Λ_3. Now, the relay sends

$$C_3 = Q_{\Lambda_3}\left(Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) + Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n) - Z_1^n - Z_2^n + Z_3^n\right),$$

at a rate

$$R_3 \leq \frac{1}{2}\log\left[1+\frac{(\beta_1+\beta_2)^2\sigma_S^2}{\sigma_{W_3}^2}+\frac{\beta_1^2(\sigma_{N_1}^2+\sigma_{W_1'}^2)+\beta_2^2(\sigma_{N_2}^2+\sigma_{W_2'}^2)}{\sigma_{W_3}^2}\right].$$

The decoder performs

$$R^n = C_3 - Z_3^n = \beta_1 S_1^n + \beta_2 S_2^n + V_1^n + V_2^n + V_3^n,$$

where V_1^n and V_2^n are defined similarly to (3.6) and

$$\begin{aligned} V_3^n =& Q_{\Lambda_3}(Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) + Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n) - Z_1^n - Z_2^n + Z_3^n) \\ & - (Q_{\Lambda_1}(\beta_1 S_1^n + Z_1^n) + Q_{\Lambda_{21}}(\beta_2 S_2^n + Z_2^n) - Z_1^n - Z_2^n + Z_3^n)). \end{aligned}$$

By choosing $\sigma_{W_k'}^2$, $k \in \{1,2\}$, such that

$$\beta_1^2\sigma_{W_1'}^2 + \beta_2^2\sigma_{W_2'}^2 + \sigma_{W_3}^2 = \beta_1^2\sigma_{W_1}^2 + \beta_2^2\sigma_{W_2}^2, \tag{3.7}$$

the linear estimate of S^n given R^n satisfies the distortion constraint D.

We compare the *compress and forward* and *compute and forward* coding strategies for the symmetric case, i.e., when $\sigma_{N_1}^2 = \sigma_{N_2}^2 = \sigma_N^2$. In such a situation, we have $\sigma_{W_1}^2 = \sigma_{W_2}^2 = \sigma_W^2$, $\sigma_{W_1'}^2 = \sigma_{W_2'}^2 = \sigma_{W'}^2$ and $\beta_1 = \beta_2 = \beta$. Therefore, it follows from (3.7) that,

$$2\beta^2\sigma_W^2 = 2\beta^2\sigma_{W'}^2 + \sigma_{W_3}^2.$$

The rate R_3 in case of *compress and forward* is smaller than *compute and forward* if

$$\begin{aligned} & 2\left(1+\frac{2\sigma_S^2+\sigma_N^2}{\sigma_W^2}\right) > 1+\frac{2\beta^2(2\sigma_S^2+\sigma_N^2+\sigma_{W'}^2)}{\sigma_{W_3}^2} \\ \Rightarrow\ & 1+\frac{2(2\sigma_S^2+\sigma_N^2)}{\sigma_W^2} > \frac{2\sigma_S^2+\sigma_N^2+\sigma_{W'}^2}{\sigma_W^2-\sigma_{W'}^2}. \end{aligned}$$

It is clear that there exist $\sigma^2_{W'} > 0$, suitably small, such that the above inequality is satisfied. It is also clear that by making $\sigma^2_{W'}$ small, we incur a corresponding increase in the sum of R_1 and R_2, above the optimal sum rate of the CEO problem.

Chapter 4

Communicating Linear Functions of Correlated Gaussian Sources over a MAC

4.1 Introduction

In this chapter, we consider the joint source channel coding problem of transmitting linear functions of two correlated Gaussians (jointly Gaussian) in a distributed fashion over an additive Gaussian noise multiple access channel (MAC). Each transmitter in the MAC has, as its message, one component of the bivariate Gaussian source and its channel input is constrained by a second moment (average power) requirement. We estimate the linear function of the two correlated sources while incurring a mean squared error at the receiver. The distortion suffered by the linear functions of the two sources is a function of the power constraints at the two transmitters as well as the channel statistics. In general, there is no separation between source and channel coding over MACs, and a joint coding scheme is desired [15].

There has been significant related work on both the source and channel aspects of this problem. In the domain of source coding, [83] considers and solves the two terminal Gaussian source coding problem. It is shown that the simple scheme of vector quantization followed by random binning is the

optimal compression scheme. This scheme is also shown to be optimal for reconstructing certain linear functions of bivariate Gaussian sources. Subsequently, a distributed lattice based coding scheme for reconstructing a linear function of jointly Gaussian sources is developed in [33]. In [81], an outer bound on the rate region for the distributed compression of linear functions of two Gaussian sources for certain correlations is presented. This bound is used to show that while existing achievable schemes are suboptimal, lattice based coding schemes are within a constant gap of the optimum for the reconstruction of certain linear functions. An improved achievable rate region for this problem is presented in [80].

In [14], the authors study the problem of lossless transmission of correlated sources over a MAC. In [34], the Gaussian lossy version of this problem is considered. It is shown that for communicating a bivariate Gaussian source over a Gaussian MAC when the receiver is interested in recovering *both* components limited by individual distortion constraints, uncoded transmission is the optimal transmission strategy below a certain signal to noise ratio (SNR) threshold. In [25], the problem of recovering a single Gaussian source observed through a Gaussian sensor network at a fusion center is solved. Uncoded transmission is exactly optimal for this problem. Lattice coding has also been previously considered for various joint source channel coding problems. In [32], the authors develop a lattice coding scheme for joint Wyner-Ziv and dirty paper coding. This idea is extended and used in [42] for the problem communicating the sum of *independent* Gaussian sources over a Gaussian MAC,

with bandwidth mismatch constraints. In [31], modulo-analog lattice codes are used for the Gaussian sensor network problem under bandwidth mismatch conditions. Thus lattice codes have been able to achieve superior performance than uncoded transmission and source channel separation for network joint source channel coding problems.

In this work, we present a lattice coding scheme for the distributed transmission of linear functions of correlated Gaussians over the MAC. The key contributions are as follows:

1. We present a lower bound on the distortion incurred while estimating linear functions of the correlated Gaussian sources over a Gaussian MAC. We consider two cases of the linear function. In one case, the lower bound is based on augmenting the receiver with a random variable that induces conditional independence between the two sources and considering a statistically equivalent system of two parallel channels from each of the transmitters to the same receiver. This genie aided bound is based on the work in [83] and [81] where the authors determine a lower bound on distortion in a source coding setting. In the other case, the lower bound is obtained by allowing the two transmitters to cooperate but constraining the transmit power of the cooperating transmitters.

2. We develop a lattice coding scheme for communicating linear functions of the two sources over this channel. The scheme we present for the MAC is similar in spirit to the scheme in [32] and [42] and is an extension to

correlated sources.

3. We show that our scheme performs "close" to the lower bound by showing that the logarithm of the ratio of the distortion achieved to the distortion lower bound is between 1 and 2 bits if the signal to noise ratio (SNR) is greater than a threshold. We also show that the lattice based transmission scheme provides an improvement over uncoded transmission in the distortion of the difference function, for correlation coefficients above a threshold.

The rest of the chapter is organized as follows. We develop the system model and notation in Section 4.2. We present a lower bound on achievable distortion in Section 4.3. In Section 4.4, we characterize the distortion achieved using an uncoded transmission scheme. In Section 4.5 we describe the lattice coding scheme and analyze its performance.

4.2 System Model and Notation

We briefly explain the notation used in this chapter before presenting the system model. We use capitals to denote random variables and boldface capitals to denote matrices. $\mathbb{E}$ is used for expectation of a random variable while we refer to an n-length vector as x^n. Throughout the chapter, logarithms used are with respect to base 2.

The system model is depicted in Fig. 4.1. Consider independent and identically distributed (i.i.d.) n-length sequences of Gaussian random vari-

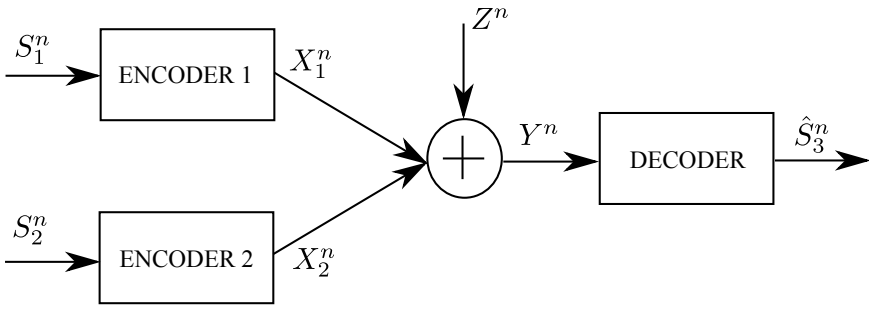


Figure 4.1: System model for joint source channel coding over MAC

ables, $\{S_{1,i}\}_{i=1}^n$ and $\{S_{2,i}\}_{i=1}^n$. The covariance matrix of $(S_{1,i}, S_{2,i})$ is given by

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

for all $i = 1, 2, ..., n$. Without loss of generality, we assume $\rho > 0$ for the purposes of this chapter. We also assume that the sources have equal variances. Both these assumptions are justified since we are interested in communicating linear functions, which can absorb deviations from these assumptions. Further, we let the linear function $S_{3,i} = S_{1,i} + cS_{2,i}$ for $i \in \{1, \ldots, n\}$. Transmitter k in the MAC has a realization of S_k^n for $k \in \{1, 2\}$. Also the number of source samples observed is equal to the number of channel uses available. Thus, the system has a bandwidth expansion factor of 1. The channel input sequence at each user is a function of the observed source sequence and is required to satisfy a power constraint. Mathematically, the channel input $X_k^n = f_k^n(S_k^n)$ for $k \in \{1, 2\}$ and the power constraint is expressed as

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(X_{k,i})^2] \leq P. \tag{4.1}$$

The noise $\{Z(i)\}_{i=1}^n$ is a sequence of i.i.d. Gaussian random variables with zero mean and variance N. The received signal at time instant i is given by

$$Y_i = X_{1,i} + X_{2,i} + Z_i.$$

We wish to estimate the sequence of the linear function $\{S_3\}_{i=1}^n$ at the receiver given the received sequence $\{Y_i\}_{i=1}^n$. The decoder applies a decoding function resulting in an estimate of the function. This is represented as $\hat{S}_3^n = \phi^n(Y^n)$. The distortion metric considered is the time average mean squared error. We now define the notion of achievability and formally state the objective of the chapter.

Definition 4.1. *We say that the tuple* (D, σ, ρ, P, N) *is* achievable *if there exist encoding and decoding functions* f_k^n, $k \in \{1, 2\}$ *and* ϕ^n *such that the distortion* D *satisfies*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(S_{3,i} - \hat{S}_{3,i})^2] \leq D.$$

The objective of the chapter is to characterize the set of achievable tuples for this problem. In other words, we wish to find the smallest D such that (D, σ, ρ, P, N) is achievable by optimizing over the encoding and decoding functions.

4.3 Lower Bound on Distortion for Linear Functions of Jointly Gaussian sources

We now present lower bounds on the distortion incurred for the distributed transmission of linear functions of correlated sources over a MAC.

Recall that the linear function $S_3 = S_1 + cS_2$ where $c \in \mathbb{R}$. We present different lower bounds on the distortion of the linear functions for two cases, as considered in the following sub-sections.

4.3.1 Case 1: $|c| \in [\rho, 1/\rho]$

Theorem 4.1. *The achievable tuple (D, σ, ρ, P, N) for communicating the linear function $S_3 = S_1 + cS_2$ when $\rho \leq |c| \leq 1/\rho$ satisfies*

$$D \geq \sup_{\lambda \in [0,1]} \frac{\sigma^2(1-\rho|c|)}{1+\frac{P}{\lambda N}} + \frac{c^2\sigma^2(1-\frac{\rho}{|c|})}{1+\frac{P}{(1-\lambda)N}}.$$

Proof. The lower bound on the distortion is obtained by augmenting the receiver with a random variable that induces conditional independence between S_1 and S_2 and splitting the MAC into two parallel channels. The idea of supplying the receiver with a random variable has been previously used in [81] to establish a bound for the problem of distributed compression of linear functions of this class. Consider the following representation for the Gaussian sources (S_1, S_2):

$$S_1 = \sqrt{\rho|c|}S + V_1$$
$$S_2 = \sqrt{\frac{\rho}{|c|}}S + V_2,$$

where S, V_1 and V_2 are independent Gaussian random variables with mean zero and variances σ^2, $\sigma^2(1-\rho|c|)$ and $\sigma^2(1-\frac{\rho}{|c|})$ respectively. Note that, by supplying the receiver with the sequence S^n, the distortion incurred can only decrease leading to a lower bound.

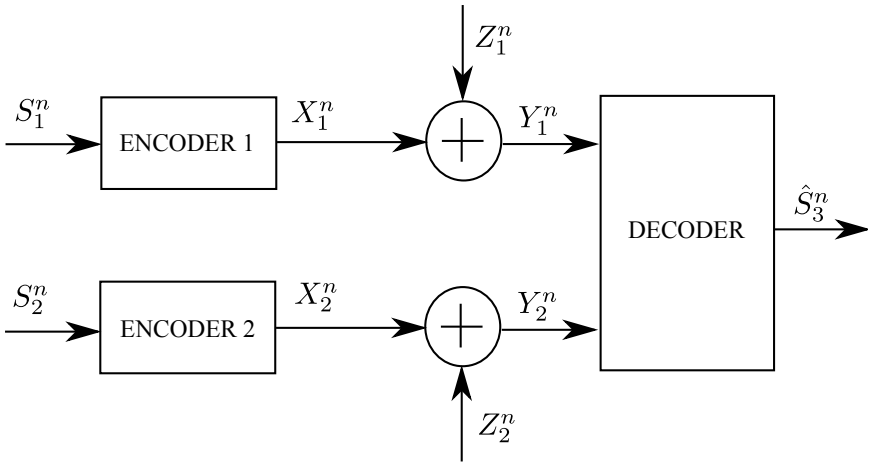


Figure 4.2: Parallel channels for lower bound on distortion

We further lower bound the distortion by considering a modified channel setting as shown in Fig. 4.2. This modified channel is a memoryless Gaussian channel which at time i is represented mathematically as

$$Y_{1,i} = X_{1,i} + Z_{1,i}$$

$$Y_{2,i} = X_{2,i} + Z_{2,i}$$

where $Z_{1,i}$ and $Z_{2,i}$ are Gaussian random variables with mean zero and variance λN and $(1-\lambda)N$ respectively, $\lambda \in [0,1]$, independent of each other and of $X_{1,i}$ and $X_{2,i}$. The receiver obtains an estimate of the linear function based on the observations of the vector (Y_1^n, Y_2^n). The distortion incurred on this channel is a lower bound on the distortion resulting from the original channel. In the original channel, the output $X_1^n + X_2^n + Z^n$ is a function of the output of the modified channel, which is the vector $(X_1^n + Z_1^n, X_2^n + Z_2^n)$. Note that the output of the original channel (in Fig. 4.1) and the sum of the outputs of

the modified channel (in Fig. 4.2) are statistically equivalent.

We consider the estimate $\hat{S}_3^n = \mathbb{E}[S_3^n|S^n, Y^n]$ since the distortion criterion is mean squared error and the conditional expectation is the optimal estimate for this criterion. Let $\hat{S}_1^n = \mathbb{E}[S_1^n|S^n, Y_1^n, Y_2^n]$ and $\hat{S}_2^n = \mathbb{E}[S_2^n|S^n, Y_1^n, Y_2^n]$. The distortion incurred in the modified channel with side information S^n at the receiver satisfies

$$\begin{aligned} D \geq & \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(S_{3,i} - \mathbb{E}[S_{3,i}|S^n, Y^n])^2] \\ \geq & \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(S_{1,i} + cS_{2,i} - \hat{S}_{1,i} - c\hat{S}_{2,i})^2] \\ = & \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(S_{1,i} - \hat{S}_1^n)^2 + (cS_{2,i} - c\hat{S}_{2,i})^2] + 2c\mathbb{E}[(S_{1,i} - \hat{S}_{1,i})(S_{2,i} - \hat{S}_{2,i})]. \end{aligned}$$

The following Markov condition

$$Y_1^n \leftrightarrow X_1^n \leftrightarrow S_1^n \leftrightarrow S^n \leftrightarrow S_2^n \leftrightarrow X_2^n \leftrightarrow Y_2^n, \tag{4.2}$$

implies that

$$\begin{aligned} \hat{S}_{1,i} = & \mathbb{E}[S_{1,i}|S^n, Y_1^n, Y_2^n] = \mathbb{E}[S_{1,i}|S^n, Y_1^n] \\ \hat{S}_{2,i} = & \mathbb{E}[S_{2,i}|S^n, Y_1^n, Y_2^n] = \mathbb{E}[S_{1,i}|S^n, Y_2^n]. \end{aligned}$$

Therefore,

$$\begin{aligned} D \geq \frac{1}{n}\sum_{i=1}^{n} \big\{ & \mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])^2] + c^2\mathbb{E}[(S_{2,i} - \mathbb{E}[S_{2,i}|S^n, Y_2^n])^2] \\ & + 2c\mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])(S_{2,i} - \mathbb{E}[S_{2,i}|S^n, Y_2^n])] \big\}. \end{aligned} \tag{4.3}$$

Now, we consider the first two terms on the RHS in (4.3) and observe that

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])^2] = \frac{\sigma^2(1-\rho|c|)}{1+\frac{P}{\lambda N}} \tag{4.4}$$

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(S_{2,i} - \mathbb{E}[S_{2,i}|S^n, Y_2^n])^2] = \frac{\sigma^2(1-\frac{\rho}{|c|})}{1+\frac{P}{(1-\lambda)N}} \tag{4.5}$$

since these are the average squared error distortions in S_1^n and S_2^n when transmitted individually over point to point Gaussian channels with noise variance λN and $(1-\lambda)N$ respectively, power constraint P, and the receiver knows S^n. The optimal distortion in (4.4) and (4.5) can be achieved using uncoded transmission [27]. In (4.4) and (4.5), we have used the fact that the conditional variance $\mathrm{Var}(S_1|S) = \mathrm{Var}(V_1) = \sigma^2(1-\rho|c|)$ and $\mathrm{Var}(S_2|S) = \mathrm{Var}(V_2) = \sigma^2(1-\frac{\rho}{|c|})$. We deal with the third term on the RHS in (4.3) as follows. We have,

$$\begin{aligned}
&\mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])(S_{2,i} - \mathbb{E}[S_{2,i}|S^n, Y_2^n])|S^n] \\
=&\mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])|S^n]\mathbb{E}[(S_{2,i} - \mathbb{E}[S_{2,i}|S^n, Y_2^n])|S^n]
\end{aligned}$$

due to the Markov condition stated in (4.2). But, by tower rule for expectations, we have

$$\mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])|S^n] = 0 \text{ a.s,}$$

for all $i = 1, 2, \ldots, n$. Thus

$$\mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])(S_{2,i} - \mathbb{E}[S_{2,i}|S^n, Y_2^n])|S^n] = 0$$

almost surely, which implies that

$$\mathbb{E}[(S_{1,i} - \mathbb{E}[S_{1,i}|S^n, Y_1^n])(S_{2,i} - \mathbb{E}[S_{2,i}|S^n, Y_2^n])] = 0. \tag{4.6}$$

Taking limsup on both sides of equation (4.3) and using (4.4), (4.5) and (4.6), for all $\lambda \in [0,1]$, we get

$$D \geq \frac{\sigma^2(1-\rho|c|)}{1+\frac{P}{\lambda N}} + \frac{c^2\sigma^2(1-\frac{\rho}{|c|})}{1+\frac{P}{(1-\lambda)N}},$$

which implies the desired bound. □

We shall denote the distortion bound described above as

$$D_1 = \sup_{\lambda\in[0,1]} \frac{\sigma^2(1-\rho|c|)}{1+\frac{P}{\lambda N}} + \frac{c^2\sigma^2(1-\frac{\rho}{|c|})}{1+\frac{P}{(1-\lambda)N}}. \tag{4.7}$$

The idea used in establishing the bound above cannot be applied for functions of type $S_1 + cS_2$ with $|c| \notin [\rho, 1/\rho]$. We obtain a lower bound for the general case in the following sub-section.

4.3.2 Case 2: $c \in \mathbb{R}$

Theorem 4.2. *The achievable tuple (D, σ, ρ, P, N) for communicating the linear function $S_3 = S_1 + cS_2$ with $c \in \mathbb{R}$ satisfies*

$$D \geq \frac{\sigma^2(1+c^2+2\rho c)}{1+\frac{2P(1+\rho)}{N}}.$$

Proof. Before we proceed to the proof of the theorem stated above, we state a lemma that is required. Without loss of generality, we can assume that $\mathbb{E}[X_{k,i}] = 0$ for $k \in \{1,2\}$ and $i = 1, \ldots, n$ [35].

Lemma 4.3. *If $\mathbb{E}[X_{k,i}] = 0$ for $k \in \{1,2\}$ and $i = 1, \ldots, n$, then*

$$\mathbb{E}[X_{1,i}X_{2,i}] \leq \rho\sqrt{\mathbb{E}[(X_{1,i})^2]}\sqrt{\mathbb{E}[(X_{2,i})^2]}.$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(X_{1,i} + X_{2,i})^2] \leq 2P(1+\rho).$$

Proof. See Lemma B.3 in [35]. □

The lower bound on the distortion in this case is obtained by allowing the two users to cooperate though the channel inputs are still power constrained as stated in (4.1). Hence, the random variables satisfy the Markov chain

$$S_3^n \leftrightarrow (X_1^n, X_2^n) \leftrightarrow Y^n \leftrightarrow \hat{S}_3^n.$$

By the data processing inequality, we have

$$I(S_3^n; \hat{S}_3^n) \leq I(X_1^n, X_2^n; Y^n). \tag{4.8}$$

By the rate distortion function for the Gaussian source S_3, we have

$$I(S_3^n; \hat{S}_3^n) \geq \frac{n}{2} \log \frac{\sigma^2(1 + c^2 + 2\rho c)}{D}. \tag{4.9}$$

Using Lemma 4.3, we get

$$I(X_1^n, X_2^n; Y^n) \leq \frac{n}{2} \log \left(1 + \frac{2P(1+\rho)}{N}\right) \tag{4.10}$$

since the output $Y^n = X_1^n + X_2^n + Z^n$ is the output of a point to point channel with input $X_1^n + X_2^n$ constrained by an average power of $2P(1+\rho)$. Combining (4.8), (4.9) and (4.10), we obtain,

$$D \geq \frac{\sigma^2(1 + c^2 + 2\rho c)}{1 + \frac{2P(1+\rho)}{N}}.$$

□

We shall use the notation,

$$D_2 = \frac{\sigma^2(1 + c^2 + 2\rho c)}{1 + \frac{2P(1+\rho)}{N}}.$$

Unlike the previous bound, this bound is valid for any $c \in \mathbb{R}$. However, for $c \in [-1/\rho, -\rho]$, we have $D_1 \geq D_2$. This can be seen as follows. Letting $\lambda = \frac{1}{2}$ in D_1 gives $\frac{\sigma^2(1+c^2+2\rho c)}{1+\frac{2P}{N}}$ as opposed to $D_2 = \frac{\sigma^2(1+c^2+2\rho c)}{1+\frac{2P(1+\rho)}{N}}$. For $c \in [\rho, 1/\rho]$, either of the bounds, D_1 or D_2 may be better depending on the problem parameters, ρ and c. In the following sections, we discuss the performance of various achievable schemes relative to the distortion bounds.

4.4 Uncoded Transmission

In this section, we compute the distortion resulting from uncoded transmission to communicate linear functions of the form $S_1 + cS_2$. Transmitter 1 sends a scaled version of the source $\sqrt{\frac{Q_1}{\sigma^2}}S_{1,i}$ at time instant i and Transmitter 2 sends $\pm\sqrt{\frac{Q_2}{\sigma^2}}S_{2,i}$ at time instant i, where $Q_1 \leq P$ and $Q_2 \leq P$. Note that the scaling is chosen such that both users satisfy their respective power constraints. Transmitter 2 chooses whether to send $\sqrt{\frac{Q_2}{\sigma^2}}S_{2,i}$ or $-\sqrt{\frac{Q_2}{\sigma^2}}S_{2,i}$ depending on c and ρ. The received sequence is given by

$$Y_i = \sqrt{\frac{Q_1}{\sigma^2}}S_{1,i} \pm \sqrt{\frac{Q_2}{\sigma^2}}S_{2,i} + Z_i.$$

The receiver determines the minimum mean squared error (MMSE) estimate of the function $S_{1,i} + cS_{2,i}$ based on the received signal Y_i.

When Transmitter 2 sends $\sqrt{\frac{Q_2}{\sigma^2}}S_{2,i}$, the distortion resulting from this process can be calculated as

$$D_{uc} = \min_{Q_1 \leq P, Q_2 \leq P} \left[\frac{\sigma^2(1-\rho^2)(c^2Q_1 + Q_2)}{Q_1 + Q_2 + 2\rho\sqrt{Q_1Q_2} + N} + \frac{2\sigma^2 c(\rho^2 - 1)\sqrt{Q_1Q_2}}{Q_1 + Q_2 + 2\rho\sqrt{Q_1Q_2} + N} + \frac{\sigma^2(1 + c^2 + 2c\rho)N}{Q_1 + Q_2 + 2\rho\sqrt{Q_1Q_2} + N} \right]. \quad (4.11)$$

For the specific case when $c = 1$, we observe that the distortion achieved by uncoded transmission for $Q_1 = Q_2 = P$,

$$D_{uc} = \frac{2\sigma^2(1+\rho)}{1 + \frac{2P(1+\rho)}{N}},$$

is exactly equal to the lower bound D_2. Thus uncoded transmission is optimal for communicating the sum of two correlated Gaussian sources. For all other cases, this uncoded scheme does not meet the lower bound for any $\rho > 0$.

When Transmitter 2 sends $-\sqrt{\frac{P}{\sigma^2}}S_{2,i}$, the MMSE estimate results in a distortion

$$D_{uc} = \min_{Q_1 \leq P, Q_2 \leq P} \left[\frac{\sigma^2(1-\rho^2)(c^2Q_1 + Q_2)}{Q_1 + Q_2 - 2\rho\sqrt{Q_1Q_2} + N} - \frac{2\sigma^2 c(\rho^2 - 1)\sqrt{Q_1Q_2}}{Q_1 + Q_2 - 2\rho\sqrt{Q_1Q_2} + N} + \frac{\sigma^2(1 + c^2 + 2c\rho)N}{Q_1 + Q_2 - 2\rho\sqrt{Q_1Q_2} + N} \right].$$

In general, the distortion resulting from this uncoded transmission scheme does not meet the lower bound for any $\rho > 0$. When $c = -1$, i.e. when we are interested in the difference of the two sources, for $Q_1 = Q_2 = P$, we get

$$D_{uc} = \frac{2\sigma^2(1-\rho)}{1 + \frac{2P(1-\rho)}{N}}. \quad (4.12)$$

For $c = -1$, we consider the lower obtained from (4.7) by substituting $\lambda = \frac{1}{2}$. This implies,

$$D_1 \geq \frac{\sigma^2(1-\rho)}{1+\frac{P}{2N}} + \frac{\sigma^2(1-\rho)}{1+\frac{P}{2N}} = \frac{2\sigma^2(1-\rho)}{1+\frac{P}{2N}}.$$

Note that when the sources are independent i.e. $\rho = 0$, the distortion achieved by uncoded transmission for $c = -1$ exactly matches the lower bound D_1. Moreover, for $\rho < \frac{1}{2}\big(1 + \frac{N}{2P}\big)$, the logarithm of the ratio of the distortion achieved by this uncoded transmission scheme to the distortion bound D_1 is less than 1 bit. Mathematically, if $\mathrm{SNR} = \frac{P}{N}$, for $\rho < \frac{1}{2}(1 + \frac{1}{2}\mathrm{SNR}^{-1})$,

$$\log \frac{D_{uc}}{D_1} \leq \log \frac{2\sigma^2(1-\rho)}{1+\frac{2P(1-\rho)}{N}} \frac{1+\frac{2P}{N}}{2\sigma^2(1-\rho)} < \log 2 = 1.$$

In the next section, we describe lattice based coding schemes which achieve within a constant of the optimal distortion and perform better than uncoded transmission for the difference function in certain correlation regimes.

4.5 Lattice Coding Scheme

We now describe a lattice scheme to communicate linear functions of the two sources. We discuss the scheme in generality and allow $c \in \mathbb{R}$. We briefly review some features of lattice codes and quantizers before we present the scheme. A lattice of dimension n is defined as the set

$$\Lambda = \{x = z\mathbf{G} : z \in \mathbb{Z}^n\}$$

where $\mathbf{G} \in \mathbb{R}^{n \times n}$ is known as the generator matrix and $\mathbb{Z}$ is the set of all integers. The quantized value of $x \in \mathbb{R}^n$ is $Q(x) = \mathrm{argmin}_{r \in \Lambda} \|x - r\|_2$. The

fundamental Voronoi region of Λ is defined as $\mathcal{V}_0 = \{x \in \mathbb{R}^n : Q(x) = 0\}$. Let V denote the volume of the Voronoi region, defined as $V = \int_{\mathcal{V}_0} dx$. The second moment of a lattice is defined as

$$\sigma^2(\Lambda) = \frac{\int_{\mathcal{V}_0} \|x\|_2^2 dx}{nV}$$

while the normalized second moment is defined as

$$G(\Lambda) = \frac{\sigma^2(\Lambda)}{V^{\frac{2}{n}}}.$$

Further, we use the notation $x \bmod \Lambda = x - Q(x)$. Suppose $x, y \in \Lambda$. The mod operation satisfies the following distributive property,

$$[(x \bmod \Lambda) + y] \bmod \Lambda = [x + y] \bmod \Lambda. \tag{4.13}$$

A lattice is said to be 'good' for channel coding if for an i.i.d. sequence of Gaussian random variables $\{Z_i\}_{i=1}^n$ with mean zero and variance $\sigma^2(\Lambda)$,

$$\lim_{n\to\infty} \Pr(Z^n \in \mathcal{V}_0) = 1.$$

The lattice is 'good' for source coding if the normalized second moment satisfies the property,

$$\lim_{n\to\infty} G(\Lambda) = \frac{1}{2\pi e}.$$

The proof existence of lattices that are simultaneously 'good' for both source as well as channel coding and their constructions are detailed in [21].

The lattice coding scheme described below is similar in nature to the lattice coding scheme used in [42] for joint source channel coding of the sum

of independent Gaussian sources. The scheme is also similar to the lattice scheme used in [32] for joint Wyner-Ziv and dirty paper coding.

Theorem 4.4. *If $\frac{P}{N} > \frac{1}{2}$, then the distortion D_{lat} achieved by the lattice coding scheme in communicating the linear function $S_1 + cS_2$ is given by*

$$D_{lat} = \frac{\sigma^2(1 + c^2 + 2\rho c)}{\frac{1}{2} + \frac{P}{N}}.$$

Proof. Consider Λ, a lattice of dimension n with second moment $\sigma^2(\Lambda) = P$. We choose the same lattice Λ at both the users such that it is 'good' for both source and channel coding. Let U_1^n and U_2^n be independent dithers (independent of each other and independent of the sources) which are uniformly distributed over the fundamental Voronoi region $\mathcal{V}_0$ and known at the receiver. The n-length channel input at each transmitter is

$$X_1^n = (\gamma S_1^n - U_1^n) \bmod \Lambda$$
$$X_2^n = (c\gamma S_2^n - U_2^n) \bmod \Lambda$$

where γ is a scalar, the choice of which is specified later. Since the dither is uniformly distributed over the fundamental Voronoi region and independent of the source, X_k^n is independent of S_l^n for $k, l \in \{1, 2\}$ [22].

The signal at the receiver is given by

$$Y^n = X_1^n + X_2^n + Z^n.$$

The decoder performs the following operations to estimate the linear function:

$$\begin{aligned} Y_1^n =& [\alpha Y^n + U_1^n + U_2^n] \bmod \Lambda \\ =& [\alpha(X_1^n + X_2^n + Z^n) + U_1^n + U_2^n] \bmod \Lambda \\ \stackrel{(a)}{=}& [\gamma(S_1^n + cS_2^n) + (\alpha - 1)(X_1^n + X_2^n) + \alpha Z^n] \bmod \Lambda \\ =& [\gamma(S_1^n + cS_2^n) + Z_1^n] \bmod \Lambda, \end{aligned}$$

where

$$Z_1^n = (\alpha - 1)(X_1^n + X_2^n) + \alpha Z^n$$

is the effective noise and (a) is due to (4.13). Note that each term in the effective noise is independent of the source since X_k^n is independent of S_l^n for $k, l \in \{1, 2\}$ and the original noise Z^n is also independent of the sources. Further, X_1^n and X_2^n are independent and by choosing $\alpha = \frac{2P}{2P+N}$, the MMSE coefficient, we reduce the variance of the effective noise to $\frac{2PN}{2P+N}$. Since Λ is chosen to be a good channel lattice, if

$$\gamma^2\sigma^2(1 + c^2 + 2\rho c) + \frac{2PN}{2P + N} \leq P, \tag{4.14}$$

we know from [22] and [32] that we can decode correctly with high probability as $n \to \infty$ and therefore

$$\lim_{n\to\infty} \Pr(Y_1^n = \gamma(S_1^n + cS_2^n) + Z_1^n) = 1.$$

Therefore, if $\frac{P}{N} > \frac{1}{2}$, we choose γ satisfying (4.14) with equality. Mathematically, γ satisfies

$$\begin{aligned} &\gamma^2\sigma^2(1 + c^2 + 2\rho c) + \frac{2PN}{2P + N} = P \\ \Rightarrow &\gamma^2\sigma^2(1 + c^2 + 2\rho c)\frac{2P + N}{2PN} + 1 = \frac{2P + N}{2N}. \end{aligned} \tag{4.15}$$

Under the assumption of correct decoding, we have

$$Y_1^n = \gamma(S_1^n + cS_2^n) + Z_1^n.$$

We construct a linear estimator of the function $S_1^n + cS_2^n$ given Y_1^n as

$$\hat{S}_3^n = \frac{\gamma\sigma^2(1 + c^2 + 2\rho c)}{\gamma^2\sigma^2(1 + c^2 + 2\rho c) + \frac{2PN}{2P+N}} Y_1^n.$$

The average distortion that is achieved is therefore given by

$$D_{lat} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[(S_{3,i} - \hat{S}_{3,i})^2] = \frac{\sigma^2(1 + c^2 + 2\rho c)}{1 + \frac{\gamma^2\sigma^2(1+c^2+2\rho c)(2P+N)}{2PN}} = \frac{\sigma^2(1 + c^2 + 2\rho c)}{\frac{P}{N} + \frac{1}{2}}. \tag{4.16}$$

where the last equality follows from (4.15). □

We now compare the distortions achieved by the lattice schemes with the lower bounds. We first consider the case when $c \in [-1/\rho, -\rho]$. The lattice based coding scheme developed above is close to the distortion bound presented in Section 4.3 in the sense that the logarithm of the ratio of the distortion bound to distortion resulting from the lattice scheme is less than one bit for any SNR $> \frac{1}{2}$. For $\lambda = \frac{1}{2}$, from (4.7), we have,

$$D_1 \geq \frac{\sigma^2(1 - \rho|c|)}{1 + \frac{P}{2N}} + \frac{c^2\sigma^2(1 - \frac{\rho}{|c|})}{1 + \frac{P}{2N}} = \frac{\sigma^2(1 + c^2 + 2\rho c)}{1 + \frac{2P}{N}}.$$

Therefore, the gap is bounded by

$$\log \frac{D_{lat}}{D_1} \leq \log \frac{\sigma^2(1 + c^2 + 2\rho c)}{\frac{P}{N} + \frac{1}{2}} \frac{1 + \frac{2P}{N}}{\sigma^2(1 + c^2 + 2\rho c)} = \log 2 = 1.$$

The SNR condition is necessary for the existence of the above lattice scheme as discussed earlier. For $c = -1$, comparing (4.12) with (4.16), we observe that

$$D_{lat} < D_{uc} \text{ for } \rho > \frac{1}{2}\left(1 + \frac{1}{2}\text{SNR}^{-1}\right).$$

Thus the lattice scheme achieves a smaller distortion in the difference than uncoded transmission for $\rho > \frac{1}{2}(1 + \frac{1}{2}\text{SNR}^{-1})$.

For the general case, when $c \in \mathbb{R}$, the logarithm of the ratio of the distortion bound to the distortion resulting from the lattice scheme is given by

$$\log \frac{D_{lat}}{D_2} = \log \frac{\sigma^2(1 + c^2 + 2\rho c)}{\frac{P}{N} + \frac{1}{2}} \frac{1 + \frac{2P(1+\rho)}{N}}{\sigma^2(1 + c^2 + 2\rho c)}.$$

Therefore for $0 \leq \rho \leq 1$ and any $\text{SNR} > \frac{1}{2}$,

$$1 \leq \log \frac{D_{lat}}{D_2} \leq 2.$$

Thus the distortion achieved by the lattice scheme is between 1 and 2 bits away from the distortion bound D_2 for any $\text{SNR} > \frac{1}{2}$. Note that, for $\text{SNR} < \frac{1}{2}$, it is possible to time share the above scheme with no transmission from either encoder, thereby resulting in an achievable scheme for the problem.

4.6 Numerical Results

We now present numerical results comparing the performance of lattice coding schemes with other schemes. In Fig. 4.3, we consider the problem of the difference of two Gaussian sources. We plot the distortion achieved by

three different coding schemes and the lower bound D_1 as a function of the correlation coefficient of the sources. The plot are generated for $\sigma^2 = 1$, $P = 0.5$ and $N = 0.05$. The three coding schemes considered are uncoded transmission, lattice coding and source channel separation. The distortion achieved by uncoded transmission and lattice codes are given by equations (4.12) and (4.16) respectively. We evaluate the distortion achieved by separation in this case using Theorem 3.1 in [33] as,

$$\begin{aligned} D = \min_{R_1,R_2} \quad & 2\sigma^2(1-\rho)(2^{-2R_1} + 2^{-2R_2}) \\ \text{such that } & 0 \leq R_1 \leq \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \\ & 0 \leq R_1 \leq \frac{1}{2}\log\left(1 + \frac{P}{N}\right) \\ & R_1 + R_2 \leq \frac{1}{2}\log\left(1 + \frac{2P}{N}\right). \end{aligned}$$

Note that lattices are used for the compression part of source channel separation when computing the difference. The distortion for separation is achieved by choosing the appropriate operating point in the capacity region of the MAC.

We see that uncoded transmission achieves a smaller distortion than other schemes at low correlation coefficients. Beyond a threshold correlation coefficient, the lattice coding scheme achieves a smaller distortion than uncoded transmission. For all correlation coefficients, the separation scheme performs very poorly. Note that while separation is optimal for recovering individual independent sources within individual distortion criteria, such a scheme performs poorly when we are interested in recovering the difference.

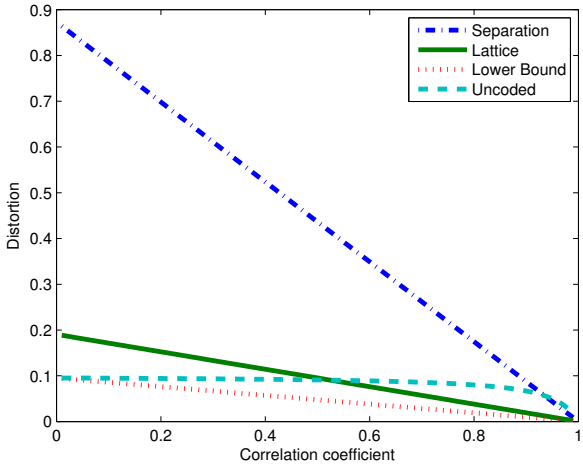


Figure 4.3: Comparison of achievable strategies and lower bound for $S_3 = S_1 - S_2$

Further, the distortion achieved by the lattice scheme is always less than two times the lower bound on the distortion. In terms of bits, the lower bound is only one bit away from the distortion achieved by the lattice scheme.

In Fig. 4.4, we consider the distributed communication of the function $S_3 = S_1 + 2S_2$ retaining the other source and channel parameters as before. The distortions achieved by uncoded transmission and lattice codes are calculated using (4.11) and (4.16). To calculate the distortion achieved by source channel

separation, we use Theorem 3.2 in [33]. Let

$$\mathcal{Q} = \Bigg\{ (q_1, q_2) : q_1 \geq 0, q_2 \geq 0$$
$$\frac{1}{2}\log\frac{(\sigma^2+q_1)(\sigma^2+q_2)-\sigma^4\rho^2}{q_1(\sigma^2+q_2)} \leq \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$
$$\frac{1}{2}\log\frac{(\sigma^2+q_1)(\sigma^2+q_2)-\sigma^4\rho^2}{q_2(\sigma^2+q_1)} \leq \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$
$$\frac{1}{2}\log\frac{(\sigma^2+q_1)(\sigma^2+q_2)-\sigma^4\rho^2}{q_1 q_2} \leq \frac{1}{2}\log\left(1+\frac{2P}{N}\right) \Bigg\}.$$

The distortion achieved by source channel separation is given by

$$D = \min_{(q_1,q_2)\in\mathcal{Q}} \left[\frac{\sigma^4 q_1(1-\rho^2)+\sigma^4 c^2 q_2(1-\rho^2)}{(\sigma^2+q_1)(\sigma^2+q_2)-\sigma^4\rho^2} + \frac{q_1 q_2 \sigma^2(1+c^2+2c\rho)}{(\sigma^2+q_1)(\sigma^2+q_2)-\sigma^4\rho^2}\right].$$

It is important to note that the Berger-Tung scheme [77] is used for compression part of source channel separation when $S_1 + 2S_2$ is desired at the destination. This is because such a scheme achieves the optimal rate distortion region for linear functions of correlated Gaussian sources [83]. The optimization described above is carried out numerically to obtain the plot in Fig. 4.4.

We see that the distortion achieved by the lattice scheme is between two and four times the lower bound on the distortion. Also, there is a threshold below which the lattice scheme achieves a smaller distortion than uncoded transmission. In general, the relative performance between uncoded transmission, separation and lattice codes, depends on the desired linear function. However, unlike uncoded transmission and separation, we are able to show

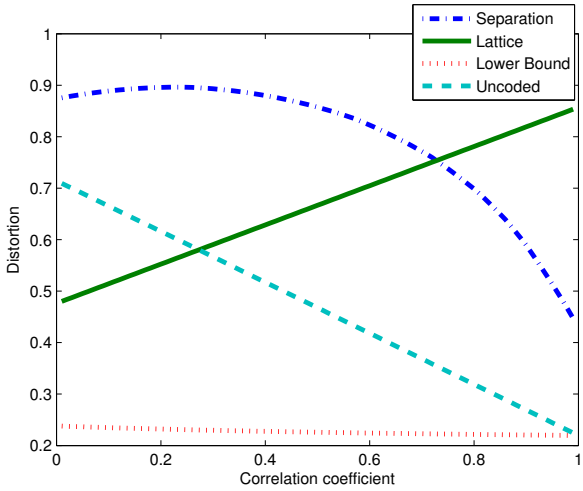


Figure 4.4: Comparison of achievable strategies and lower bound for $S_3 = S_1 + 2S_2$

that the distortion achieved by the lattice scheme always lies within a constant factor of the optimum.

Chapter 5

Reduced Reference Entropic Differencing for Image Quality Assessment

5.1 Introduction

Reduced reference (RR) quality assessment (QA) algorithms involve sending or supplying some amount of information about the reference along with the distorted image that is useful in quality computation. For example, the concept of *quality aware* images was proposed in [93], where partial reference image information is embedded within the image and can be extracted reliably despite distortions. The information embedded could for example, be the statistical parameters of the distribution of wavelet coefficients obtained by a multi scale-space-orientation decomposition of the reference image. Two parameters of a generalized Gaussian distribution and the error in approximating the empirical coefficients by this distribution are transmitted for every subband. The quality is based on computing the Kullback-Leibler (KL) divergence between the parametrized distribution of the reference and the empirical distribution of the distorted image. The performance of this algorithm is good only for certain individual distortion categories. This idea is further extended in [37], where an additional divisive normalization transform step is introduced before computing the KL divergence to improve performance. However,

this algorithm depends on a number of parameters that need to be trained on databases. There has also been prior work on non-statistical reduced reference methods that compute a generalized norm between selected wavelet coefficients of the reference and distorted images [39]. The coefficients are selected in a way that reduces the information required while still maintaining good performance.

Algorithms based on multiscale geometric analysis including curvelets, bandlets, wavelets and contourlets are developed in [24]. These algorithms depend on parameters that need to be tuned delicately on different databases. Moreover, their performances degrade severely with the reduction in the data rate required from the reference. Algorithms designed in [8] for color images, perform well when tested on images belonging to certain distortion classes such as JPEG or JPEG2000. While distributed source coding ideas are applied to approximate mean squared error in [12], RR QA algorithms based on color distribution of images are developed in [52], and training based approaches are used in [20]. These algorithms are either limited in their ability to achieve good performance across different distortion types or involve training on databases. The algorithm by [98] based on Weibull statistics of wavelet coefficients achieves good performance at a given data rate. However, what we desire in this work, is a family of algorithms that achieve graceful degradation in performance with the reduction in data rate and the ability to achieve better performance with increase in data rate.

In this chapter, we develop a new framework of reduced reference QA

algorithms that are information theoretic. We consider natural image approximations of the distorted image in the sense that wavelet coefficients of distorted images will be fitted with Gaussian scale mixture distributions. In effect, this dissertation approaches the problem of quality assessment from the perspective of measuring distances between the reference image and the projection of the distorted image on to the space of natural images. The algorithms compute the average difference between scaled entropies of wavelet coefficients of reference and projected distorted images that are obtained at the output of a neural noise channel. A family of algorithms are proposed depending on the subband in which the quality computation is carried out and the amount of information required from the reference image. This framework allows us to study how the performance of these information theoretic RR QA algorithms decays with reduction in the amount of information used from the reference. Since, the reference information scales with the size of the images, the algorithms are also applicable in scalable image and video QA [59]. Further, the algorithms allow for bidirectional computation of quality, by which we mean that the quality of the distorted image can also be computed at the reference if relevant information from the distorted image is made available. In other words, by supplying reduced information from the distorted image through a feedback channel, its quality can be computed at the reference without sending any information in the forward channel. This feature has potential applications in image/video quality monitoring in networks, which requires feeding back the quality at different nodes in the network to the sender. Of course,

since the quality index is based on the absolute difference, it does not indicate which of the two images is the reference image. In nearly any imaginable scenario, we know which of the two images is the reference. Another interesting feature of these algorithms is that they are not dependent on any parameters that need to be trained on databases. Depending on the bandwidth available for information supplied either from the reference or distorted image, one algorithm from this class may be picked for desired applications. The framework also allows users to choose an algorithm from this class for general purpose or distortion specific quality assessment.

The rest of the chapter is structured as follows. In the following section, we describe prior information theoretic approaches that motivate the algorithms developed in this dissertation. In Section 5.3, we present the main theory and resulting algorithms of the chapter. We discuss perceptual interpretations of the algorithms in Section 5.4, and present numerical results in Section 5.5. It is found that the performance attained is highly competitive on a large and comprehensive database of distorted images and subjective scores.

5.2 Information theoretic approaches to QA

Information theoretic methods for quality assessment have produced some of the best performing *full reference* (FR) QA algorithms, including the information fidelity criterion (IFC) [65] and the visual information fidelity (VIF) index [63]. The motivation for such an approach is that image distortions tend to disturb the natural statistics of images and quantifying this

disturbance can determine quality. It is also based on the assumption that such modifications of natural image statistics are perceptually noticeable. Mathematically, IFC and VIF compute the amount of mutual information shared between the reference and distorted images in the wavelet domain under a natural scene statistic (NSS) model. Moreover, both these algorithms possess psychovisual properties that are desirable for quality assessment. In particular, a number of similarities are drawn between these information theoretic methods and perceptual properties such as masking, suprathreshold effects, error pooling, scale-space-orientation decomposition, and so on, that make these algorithms perceptually and statistically appealing.

The mutual information terms in IFC and VIF can be shown to be functions of the correlation coefficients between patches of wavelet coefficients under the assumed NSS model, which is a Gaussian scale mixture (GSM) model [60]. This means that the entire reference image is required in order to compute the IFC/VIF indices (or the correlation coefficients). Hence these are both full reference QA algorithms and not useful in no reference and reduced reference scenarios. This leads to the question: what information theoretic quantities can we compute using only reduced or no information from the reference?

The most successful general purpose reduced reference quality metrics are based on computing the KL divergence between the reference and test images [93]. A divisive normalization transform (DNT) step before computation of the KL divergence was shown to further improve performance [37]. The

DNT resembles a divisive normalization stage used in visual neural models and also allows for local processing before computing the KL divergence, which is a global measure of distances. It is interesting to note that while the quality indices based on shared mutual information have perceptual interpretations such as masking, suprathreshold effects, error pooling, and so on [65], those based on KL divergence by themselves do not possess these features. Such psychovisual properties have to be additionally introduced, for example, by the divisive normalization transform as in [37] to improve performance. This leads to the question of whether we can design information theoretic QA indices in reduced reference scenarios that possess other desirable psychovisual properties in order to further improve their performance.

As a possible answer to questions posed in this section, the approach that we take in this dissertation, is to compute the average difference of scaled local entropies in the wavelet domain between the reference and the distorted image. In essence, the quality indices proposed compute the amount of local information difference between the reference and distorted images in a subband. Further, this quantity can be computed in a distributed fashion between the reference and distorted images, making it a reduced reference algorithm. The reason we compute the difference of entropies is primarily due to the constraint of the problem (reduced reference) than by choice. However, we show later on that this procedure still possesses desirable psychovisual properties leading to excellent performance of the algorithms. We present the details of the algorithms in the following section, reserving the perceptual interpretation of the

algorithms for a later section.

5.3 RR QA Algorithms

We now describe the main theory on which the algorithms proposed in the chapter are based. The source model considered here closely follow the assumptions in [65] while we approximate the wavelet coefficients of the distorted image to also follow Gaussian scale mixture distributions. The resulting RR algorithms utilize the wavelet coefficients obtained by a steerable pyramid decomposition of the reference and distorted image into subbands at different orientations and scales [67]. Let K be the total number of subbands obtained as a result of this decomposition. The wavelet coefficients in subband k, $k \in \{1, 2, \ldots, K\}$, are partitioned into M_k non-overlapping blocks, each block containing N coefficients of size $\sqrt{N} \times \sqrt{N}$. Non-overlapping blocks are assumed independent and identically distributed (i.i.d.). Although, wavelet coefficients in adjacent blocks in scale, space or orientation may not be independent, we make this assumption in order to simplify the quality index.

5.3.1 Source Model

Wavelet coefficients of natural images are modeled well by Gaussian scale mixture distributions. Such models have also proved very useful in various image processing applications including quality assessment [63], denoising [50] and so on. Let $\bar{C}_{mk} = (C_{1mk}, C_{2mk}, \ldots, C_{Nmk})$ denote the vector of coefficients in block m, $m \in \{1, 2, \ldots, M_k\}$ of subband k of the reference image.

Therefore,

$$\bar{C}_{mk} = S_{mk}\bar{U}_{mk},$$

where $\bar{U}_{mk} \sim \mathcal{N}(0, \mathbf{K}_{U_k})$ and S_{mk} is a scalar random variable that modulates the covariance matrix of the block $\bar{C}_{mk}$. Also, S_{mk} and $\bar{U}_{mk}$ are independent. Subband k is associated with the covariance matrix $\mathbf{K}_{U_k}$ and $\bar{U}_{mk}$ for each block m, is distributed identically. Thus, the wavelet coefficient block $\bar{C}_{mk}$, when conditioned on the realization $S_{mk} = s_{mk}$, is distributed according to a Gaussian model with a covariance matrix $s_{mk}^2\mathbf{K}_{U_k}$. Further, S_{mk} and $\bar{U}_{mk}$ are each independent over m and k.

5.3.2 Distortion Model

Distortions introduced in a natural image may take it outside the space of natural images. As a result, it is possible that the wavelet coefficients of these images do not follow a Gaussian scale mixture distribution. We approach the problem of quality assessment by projecting the distorted image on to the space of natural images. This means that we model the wavelet coefficients of the distorted image as well as a Gaussian scale mixture distribution. If the distortion process were to retain the distorted image within the space of natural images, we would not make any error by employing such a model. We measure quality as a distance between the reference and a natural image approximation of the distorted image. We show later that the approach of approximating the wavelet coefficients of a distorted image by a GSM model results in RR QA algorithms that perform very well in predicting the quality

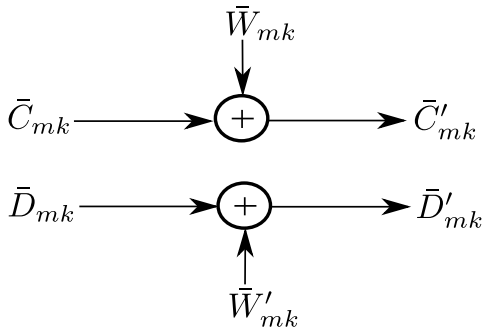


Figure 5.1: System model for quality assessment

scores of images. Denote $\bar{D}_{mk} = (D_{1mk}, D_{2mk}, \ldots, D_{Nmk})$ as the vector of coefficients in block $m \in \{1, 2, \ldots, M_k\}$ of subband k of the distorted image. We have

$$\bar{D}_{mk} = T_{mk}\bar{V}_{mk},$$

where $\bar{V}_{mk} \sim \mathcal{N}(0, \mathbf{K}_{V_k})$ and T_{mk} is the scalar premultiplier random variable as in the reference image. The independence assumptions are similar to the reference image. We now describe the RR index.

5.3.3 RR Quality Index

We additionally model the perceived reference and distorted images as passing through an additive neural noise channel, where the noise is assumed to be a zero mean Gaussian random vector for each block of coefficients. The neural noise model accounts for uncertainty introduced by neural processing of the visual signal [63]. The resulting system model is shown in Fig. 5.1. We

have

$$\bar{C}'_{mk} = \bar{C}_{mk} + W_{mk}$$
$$\bar{D}'_{mk} = \bar{D}_{mk} + W'_{mk},$$

where $W_{mk} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$ and $W'_{mk} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$, W_{mk} and W'_{mk} are independent of each other, independent of $\bar{C}_{mk}$ and $\bar{D}_{mk}$ and independent across the indices m and k. The reduced reference quality indices that we introduce and which we term Reduced Reference Entropic Differences (RRED) indices are defined as the average of the absolute value of the difference between the scaled entropies of the neural noisy reference and distorted images, conditioned on the realizations of the respective premultiplier random variables in a subband.

Let the eigen values of $\mathbf{K}_{U_k}$ be $\alpha_{1k}, \alpha_{2k}, \ldots, \alpha_{Nk}$ and the eigen values of $\mathbf{K}_{V_k}$ be $\beta_{1k}, \beta_{2k}, \ldots, \beta_{Nk}$. In the following, assume that $\mathbf{K}_{U_k}$ and $\mathbf{K}_{V_k}$ are full rank matrices. If this is not true, then the index is calculated by using the positive eigen values alone. The entropy of a reference image block m in subband k conditioned on $S_{mk} = s_{mk}$ is given by [1]

$$\begin{aligned} h(\bar{C}'_{mk}|S_{mk} = s_{mk}) &= \frac{1}{2} \log \left[(2\pi e)^N (s_{mk}^2 |\mathbf{K}_{U_k}| + \sigma_W^2 \mathbf{I}_N) \right] \\ &= \sum_{n=1}^{N} \frac{1}{2} \log \left[(2\pi e)(s_{mk}^2 \alpha_{nk} + \sigma_W^2) \right]. \end{aligned}$$

[1] All logarithms in the chapter are with respect to base 2.

Similarly, the entropy of the distorted image block conditioned on $T_{mk} = t_{mk}$ is given by

$$h(\bar{D}'_{mk}|\bar{T}_{mk} = t_{m,k}) = \sum_{n=1}^{N} \frac{1}{2} \log \left[(2\pi e)(t_{mk}^2 \beta_{nk} + \sigma_W^2) \right]. \tag{5.1}$$

Define scaling factors,

$$\gamma_{mk}^r = \log(1 + s_{mk}^2) \text{ and } \gamma_{mk}^d = \log(1 + t_{mk}^2).$$

The entropies conditioned on the realizations of the premultiplier random variables are multiplied by the above scalars before computing the difference. These scalars are increasing functions of the premultiplier random variables, that tend to zero as the premultiplier tends to zero and saturate at high values. The imposition of these scalars before computing the difference has many advantages. They lend a local character to the algorithm imposing additional local effects on the entropy terms. The other benefit of these weights is in the context of those algorithms that operate with extremely small neural noise variance. In such a setting, these help saturate the entropy terms at locations having extremely small premultiplier random variable realizations, i.e. those that are close to zero. This helps avoid numerical instabilities in the computation of the entropy differences, especially when computing the logarithm of very small variance values.

We present a family of algorithms, by varying the subband in which quality is evaluated and the amount of information that is required from each subband for quality computation. First, we discuss algorithms obtained by

varying the subband in which quality computation is carried out, alone. In these algorithms, the scaled entropies at each block in one particular subband k, $\{\gamma^r_{mk} h(\bar{C}'_{mk}|S_{mk} = s_{mk})\}_{m=1}^{M_k}$ of the reference image are used to evaluate quality. Since different subbands have different sizes, the number of blocks reduces from the subbands at the finest to the subbands at the coarsest scales of the wavelet decomposition. Thus, the number of weighted entropy terms required is equal to the number of blocks in the corresponding subband, M_k.

The reduced reference QA index corresponding to subband k, when M_k scalars are available from the reference is given by

$$\mathrm{RRED}_k^{M_k} = \frac{1}{L_k} \sum_{m=1}^{M_k} |\gamma^r_{mk} h(\bar{C}'_{mk}|S_{mk} = s_{mk}) - \gamma^d_{mk} h(\bar{D}'_{mk}|T_{mk} = t_{mk})|,$$

where L_k is the size (number of coefficients) of the subband k. The above index is a reduced reference index since M_k is less than the size of the image. We require all the M_k entropy terms since the absolute values of the differences are summed up. Note that the maximum size of M_k over all subbands is equal to the size of the image divided by N. Also, the information required reduces when the quality is evaluated in the coarser bands as M_k reduces from subbands at finer scales to coarser scales. Either image can compute the index using the entropy information from the other image. The absolute value of the difference is calculated, since the nature of the distortion process could lead to either an increase or a decrease in entropy. We only wish to measure the magnitude of the difference to evaluate quality. This also implies that the RRED indices are always positive. Also, any enhanced image would

show a difference in entropies and the difference should be interpreted as an improvement in quality.

The amount of information required from a subband can also be reduced by summing scaled entropy terms over patches and sending the sum of these scaled entropies instead of all the entropy terms. This is equivalent to filtering the image of weighted entropies in a subband using rectangular windows of sizes $b \times b$ and subsampling by b in each dimension, where b is a natural number that represents the size of the patches. This procedure results in loss of performance with subsampling as illustrated in Section 5.5. Let Λ_k denote the number of subsampled blocks and let $\lambda \in \{1, 2, \ldots, \Lambda_k\}$ index the block. Every $m \in \{1, 2, \ldots, M_k\}$ belongs to one of the subsampled blocks $B_{\lambda k}$ in subband k. Define

$$g^r_{\lambda k} = \sum_{m \in B_{\lambda k}} \gamma^r_{mk} h(\bar{C}'_{mk} | S_{mk} = s_{mk}) \text{ and}$$
$$g^d_{\lambda k} = \sum_{m \in B_{\lambda k}} \gamma^d_{mk} h(\bar{D}'_{mk} | T_{mk} = t_{mk}).$$

Then the RR quality index in subband k when Λ_k scalars are available from the reference is given by

$$\text{RRED}^{\Lambda_k}_k = \frac{1}{L_k} \sum_{\lambda=1}^{\Lambda_k} |g^r_{\lambda k} - g^d_{\lambda k}|.$$

Thus, by filtering and subsampling, we can reduce the information required from every subband for quality computation. For example, for $b = 2$, if the subband is filtered by windows of size 2×2 and subsampled by a factor

of 2 in each dimension, then the number of entropy terms required reduces from M_k to $\Lambda_k = M_k/4$. This is another method of reducing the amount of information required as against evaluating quality in subbands at coarser scales. RRED_k^1 denotes the algorithm in which all the scaled entropy terms in the subband are added and only the sum, which is a single scalar, is required for quality computation. Since only a single number is needed, this may be considered as an almost reference free algorithm. Here, we do not imply that the almost reference free algorithm approximates a no reference algorithm, but rather that just a single number is required. The algorithm still requires the single number without which quality cannot be computed.

The two methods described above illustrate how the amount of information can be reduced gradually from an almost full reference scenario to an almost reference free scenario. Moreover, the filtering and subsampling procedure can be performed in coarser bands to further reduce the information. This results in a family of algorithms with varying performance levels. There are other variations that could be performed on the class of algorithms discussed so far. Let $\mathcal{K}$ denote the set of subbands from which quality indices are combined. For example, the quality index could be given as $\sum_{k\in\mathcal{K}} \mu_k \text{RRED}_k^b$ where μ_k are scalars used to weight the respective bands differently. We evaluate one such weighting strategy in Section 5.5.

5.3.4 Estimation of Parameters

In order to compute the QA index, it is necessary to estimate $\bar{s}$, $\bar{t}$, $\mathbf{K}_{U_k}$ and $\mathbf{K}_{V_k}$, $k \in \{1, 2, \ldots, K\}$. The procedure outlined here is similar to the estimation of reference image parameters in [64]. We obtain maximum likelihood estimates for the above parameters. Without loss of generality, assume $\sum_{m=1}^{M_k} s_{mk}^2 = 1$ and $\sum_{m=1}^{M_k} t_{mk}^2 = 1$. Therefore, the ML estimates of $\mathbf{K}_{U_k}$ and $\mathbf{K}_{V_k}$ are given by

$$\hat{\mathbf{K}}_{U_k} = \sum_{m=1}^{M_k} \frac{\bar{C}_{mk}\bar{C}_{mk}^T}{M_k} \quad \hat{\mathbf{K}}_{V_k} = \sum_{m=1}^{M_k} \frac{\bar{D}_{mk}\bar{D}_{mk}^T}{M_k}.$$

Since the wavelet coefficients are conditionally Gaussian distributed, the ML estimates of s_{mk}^2 and t_{mk}^2 are given by

$$\hat{s}_{mk}^2 = \frac{\bar{C}_{mk}^T \mathbf{K}_{U_k}^{-1} \bar{C}_{mk}}{N} \quad \hat{t}_{mk}^2 = \frac{\bar{D}_{mk}^T \mathbf{K}_{V_k}^{-1} \bar{D}_{mk}}{N},$$

for $m \in \{1, 2, \ldots, M_k\}$.

5.4 Perceptual interpretation of the algorithms

We now cast the RR QA indices proposed in the previous section against perceptual principles often used in developing QA algorithms [74].

5.4.1 Scale-Space-Orientation Decomposition

The first step in the RR algorithm, similar to IFC/VIF, is a wavelet decomposition of the image at multiple scales and orientations. This step imitates the signal processing that happens in the primary visual cortex (area V1)

of the human visual system. There are different wavelet transforms that could be used for the multi-scale multi-orientation decomposition. The Gabor family of wavelets are widely used in image processing owing to the fact that neural responses in the primary visual cortex are well modeled by Gabor filters [17]. Further, the Gabor functions achieve the lower bound on the uncertainty in space and spatial frequency and are thus simultaneously localizable. Successful video quality assessment algorithms such as the MOVIE index [61] employ Gabor filters to decompose video data prior to performing quality assessment.

On the other hand, orthogonal wavelet transforms used in multiscale wavelet analysis suffer from aliasing or lack of shift invariance. A small shift in space in the image signal could lead to a significant change in the subband responses. Steerable pyramid is an alternate multiscale multiorientation decomposition that has the advantage of being shift, scale and rotation invariant. Further, steerable pyramids have been successfully deployed in a variety of image processing applications including compression, denoising, deblurring, and so on. In the context of image quality assessment, full reference image quality indices such as IFC and VIF use the steerable pyramid wavelet decomposition. The quality indices are evaluated in the wavelet domain, using a Gaussian scale mixture distribution (GSM) model on the wavelet coefficients to compute information theoretic quantities. Likewise, we use the steerable pyramid decomposition to accomplish reduced reference image quality assessment.

5.4.2 Contrast Masking

The contrast masking principle refers to the phenomenon by which the visibility of distortions in a signal component is inhibited by the presence of a masker having a similar orientation or scale at a given location. Contrast masking has been modeled in various ways in the literature, the two main methods being threshold elevation [36, 44] and gain control through divisive normalization [23, 57, 94].

Divisive normalization refers to models of perceptual processing, whereby neural responses are divided by local (neural responses adjacent to the given location in space, orientation and scale) energy in the responses. Our approach to entropic RR QA algorithm design uses conditioning on the realizations of the premultiplier random variables, which is analogous to divisive normalization processes in the primary visual cortex (area V1) of the human visual system. Both divisive normalization and conditioning seek to reduce the amount of dependence in local blocks of wavelet coefficients or neural responses at the output of a multichannel decomposition. We use the notation from the previous section where a block of wavelet coefficients $\bar{C} = S\bar{U}$. While dividing by S results in the vector $\bar{U}$ and the dependence structure within it, conditioning on S statistically achieves the same objective, leaving us with just the dependence in $\bar{U}$. Therefore, divisive normalization approximates conditioning and vice versa. Reduction in the dependence of local responses is also an important perceptual phenomenon in the human visual system which can be accomplished by both divisive normalization and conditioning.

The difference of weighted entropies also ensures numerical stability of the distortions at locations where the energy of the coefficients is extremely small, especially when the neural noise variance is also small. The weights at these locations tend towards zero, thereby saturating the difference between the reference and the distorted images at these locations. This effect is similar to the role that saturation constants perform in the divisive normalization of errors [74]. We hypothesize that the neural noise variance is also related to the saturation constants in contrast sensitivity models. While saturation constants achieve numerical stability in areas of low signal energy, they also saturate the responses to zero in such areas. The neural noise variance precisely achieves this while computing the local entropy terms. These ensure that the local entropy terms are stable yet small in areas of low signal energy.

5.4.3 Suprathreshold Effects

Suprathreshold effect refers to the phenomenon by which distortions are perceivable only if they are at or above a threshold distortion level. Further, variations in the level of suprathreshold distortions decreases as the degree of distortion increases. The logarithm operation in the algorithm essentially accomplishes this desirable property. Note that the entropy of Gaussian random variables is expressed by the logarithm of the variance. Although we compute the logarithm of local variances of the reference and distorted image coefficients, scale them appropriately and then compute the difference, this is equivalent to computing the logarithm of the ratio of variances raised to

powers (corresponding to the weights). Thus the ratio of variances raised to powers is the underlying distortion measure on which a nonlinear logarithmic function is applied.

5.4.4 Error Pooling

The error pooling strategy in these algorithms is a two step process depending on the amount of reference information on which the algorithms operate. The algorithms which send all the entropy terms from the reference image use a single pooling mechanism in which the absolute values of the differences of scaled entropies are averaged in the last step. The single number algorithms also follow a single pooling strategy where all the scaled entropy terms are first averaged and then the absolute values of these differences are calculated. In between the two extremes, the algorithms use a two step pooling strategy where some local entropy terms are pooled first and another pooling is performed on the absolute differences of the pooled entropy terms of the first stage. Pooling strategies in both stages can be thought of as a Minkowski error pooling strategy with exponent 1. The Minkowski pooling strategy is popularly used in human visual system based approaches to image quality assessment such as [74], although there is no evidence to support that such a pooling method is a good model for the aggregation of information in visual area MT (middle temporal) of the human visual system.

5.5 Results and Discussion

We conducted experiments on the two largest and best-known image quality databases: LIVE Image Quality Assessment Database [66] and the Tampere Image Database [49] of distorted images and perceptual scores. The LIVE database contains five different distortion categories including JPEG2000, JPEG, additive white Gaussian noise, Gaussian blur and bit errors due to transmission of JPEG2000 images over a fast fading channel. There are a total of 779 distorted images across all distortion categories. Note that the reduced reference algorithm operates without knowledge of the distortion type. We present the results of various algorithms belonging to the framework described in this chapter. One class of algorithms uses all of the entropy terms in every subband to compute quality. In another class of algorithms, the weighted entropies of every subband are filtered and subsampled at different rates, producing different algorithms for the same subband.

Both the reference and distorted images are decomposed into different subbands using a steerable pyramid wavelet decomposition using 6 orientations at 4 scales [67]. Thus there are a total of 26 subbands in the wavelet decomposition. The algorithm was implemented using blocks of size 3×3 in each subband, implying a value of $N = 9$. However, the algorithms are robust to the exact choice of the size of neighborhoods and we get approximately the same performance for different sizes.

In Table 5.1, we show the performance results of the algorithm obtained by computing $RRED_k^{M_k}$ for all the vertical oriented subbands at different

scales. The analysis in [65] suggests that human subjects are more sensitive to horizontal and vertical orientations than others. Further, we observe that the performance obtained by choosing the vertical subbands is marginally better than the horizontal subbands. The vertically oriented subbands are indexed by $k = 4, 10, 16, 22$ from the coarsest to the finest scale, i.e. at levels 1 through 4. We choose $\sigma_W^2 = 0.1$. Note that this is the same value of neural noise variance that has been used in prior algorithms [63] and we do not imply any training on the database. The performance of the algorithms in the RRED framework has proved to be robust to the choice of the neural noise variance, especially in the regime of high information from the reference.

Table 5.1: SROCC between RRED indices at different scales and LIVE Image Quality Assessment Database scores. JP2K - JPEG2000, NOS - No. of scalars, FF - Fast fading errors

Index	JP2K	JPEG	AWGN	Blur	FF	Overall	NOS
$\text{RRED}_4^{M_4}$	0.9536	0.9772	0.9763	0.9221	0.7549	0.8964	L/576
$\text{RRED}_{10}^{M_{10}}$	0.9631	0.9777	0.9769	0.9595	0.8523	0.9343	L/144
$\text{RRED}_{16}^{M_{16}}$	0.9580	0.9759	0.9780	0.9678	0.9427	0.9429	L/36
$\text{RRED}_{22}^{M_{22}}$	0.9363	0.9405	0.9779	0.9236	0.9377	0.9149	L/9
PSNR	0.8951	0.8812	0.9853	0.7812	0.8904	0.8754	L
VIF[63]	0.9696	0.9846	0.9858	0.9728	0.9650	0.9636	L

In Table 5.1, the Spearman rank order correlation coefficient (SROCC) between the scores of one class of the RRED indices and subjective (DMOS) scores from the LIVE image database are shown. The SROCC helps analyze how well the prediction monotonicity is preserved between the subjective ratings of quality and the scores output by the RR QA algorithm. The performance is shown for each distortion as well as over all on the database. A

comparison is also drawn between the amount of information required from the reference for computation of the quality indices for the respective algorithms. Throughout this section, L refers to the size of the image in pixels. The row labeled 'No. of scalars' indicates the amount of information required from the reference for quality computation. We also included the performance of FR QA algorithms such as PSNR and VIF for comparison. One important observation is that some of the RRED indices *perform nearly as well as the best performing FR QA algorithms* such as VIF. Further, most of the RRED indices considerably outperform PSNR for all distortion types (except images distorted by Gaussian noise) as well as on the entire database. Even for images distorted by Gaussian noise, the performance of the RRED indices is comparable with that of PSNR, which is an FR algorithm. The variation in performance of the RRED indices at different scales reveals that even though evaluating the quality in a coarser subband needs less information, it could potentially outperform the algorithm evaluated in a finer scale using more information. This suggests that transmitting the right information is crucial to obtain high performance QA algorithms.

The correlation analysis reveals that subbands at certain scales are more sensitive to particular distortions or artifacts. While blur and fast fading errors are better captured by subbands at finer scales, the coarser scales perform better for compressed and noisy images. This is because edge like artifacts introduced by these distortions are reflected better in the coarser subbands. The performance of fast fading errors goes down with the fineness of the scale

even though the overall performance on the database is still very good. Given an application, a user could evaluate the algorithm in a subband or subbands that yield the best performance.

It is worth mentioning that an understanding of how subjective scores change with changes in the viewing distance or the size of the images is very limited. The work in [5] represents a subjective study on viewing distance conditions for JPEG and JPEG2000 images, while the work in [26] analyzes the effect of downsampling the image, prior to objective quality assessment. We remark that it is indeed possible that depending on the viewing conditions (in particular, the distance at which images are viewed), RRED evaluated at different scales could be more perceptually relevant.

We show the linear correlation coefficient (LCC) scores obtained between DMOS (difference of mean opinion scores) or subjective scores of quality and the RRED indices in Table 5.2. A logistic nonlinearity is applied to the RRED indices before computing the linear correlation between the quality scores of the algorithm and the subjective scores available with the database. The nonlinearity relation is described by

$$\begin{aligned} \text{Quality}(x) &= \beta_1 \,\text{logistic}(\beta_2, x - \beta_3) + \beta_4 x + \beta_5 \\ \text{logistic}(\tau, x) &= \frac{1}{2} - \frac{1}{1 + \exp(\tau x)}. \end{aligned}$$

We observe the same trends in Table 5.2 as in Table 5.1 for SROCC. We also show the logistic fit of scores of the objective algorithms listed in the table in Fig. 5.2. The flattening of the curves from (a) to (d) indicates that the RRED

indices at coarser scales tend to cluster the quality indices of most images close to zero. As we move from (a) to (d), the RRED indices are now evaluated in finer scales, which provide a better separation of quality scores. The outliers in (a), (b) and (c) correspond to images containing a large area of high frequency textured regions. Since (a), (b) and (c) are evaluations of the RRED indices at coarser scales, they do not accurately capture the degradation in quality in these high frequency regions leading to deviations in the logistic fit.

Table 5.2: LCC between RR algorithms at different scales and LIVE Image Quality Assessment Database scores. JP2K - JPEG2000, NOS - No. of scalars, FF - Fast fading errors

Index	JP2K	JPEG	AWGN	Blur	FF	Overall	NOS
$\text{RRED}_4^{M_4}$	0.9600	0.9819	0.9813	0.9318	0.7838	0.9066	L/576
$\text{RRED}_{10}^{M_{10}}$	0.9688	0.9820	0.9838	0.9661	0.8688	0.9349	L/144
$\text{RRED}_{16}^{M_{16}}$	0.9629	0.9793	0.9845	0.9698	0.9413	0.9385	L/36
$\text{RRED}_{22}^{M_{22}}$	0.9401	0.9457	0.9682	0.8980	0.9220	0.9099	L/9
PSNR	0.8995	0.8899	0.9861	0.7837	0.8897	0.8723	L
VIF [63]	0.9476	0.9873	0.9883	0.9745	0.9696	0.9604	L

The algorithm evaluated in subband 10 yields excellent overall performance (in terms of both SROCC and LCC) as well as for compression, noise and blur distortions. The only drawback of the algorithm is the reduced performance in the category of images corrupted by fast fading errors. The overall performance decays in the subband at the finest scale while the performance in the individual categories is still very good. Also, we observe that the performance of the subband at the second finest scale (subband 16) is uniformly very good for all distortion types and on the overall database. We now study the

effect of filtering and subsampling the scaled entropies on subband 16. Similar trends are observed for the other subbands as well.

Table 5.3: Effect of filtering and subsampling on RR algorithms - SROCC. Index 1-RRED$_{16}^{M_{16}}$, 2-RRED$_{16}^{M_{16}/4}$, 3-RRED$_{16}^{M_{16}/16}$, 4-RRED$_{16}^{M_{16}/64}$, 5-RRED$_{16}^{M_{16}/256}$.

Index	JP2K	JPEG	AWGN	Blur	FF	Overall	NOS
1	0.9580	0.9759	0.9780	0.9675	0.9427	0.9429	L/36
2	0.9600	0.9760	0.9748	0.9429	0.9380	0.9359	L/144
3	0.9611	0.9766	0.9688	0.9674	0.9338	0.9169	L/576
4	0.9580	0.9726	0.9626	0.9679	0.9275	0.8865	L/2304
5	0.9490	0.9623	0.9502	0.9655	0.9202	0.8543	L/9296

Table 5.3 clearly demonstrates the degradation in the performance of the RR algorithms with subsampling. Observe that even though the overall correlations are reduced by increases in the degree of subsampling, the correlation score for each individual distortion category remains very competitive with state of the art FR algorithms. The reduced performance on the overall database can be attributed to the different ranges of quality scores that occur for different distortion types. Certain distortions such as JPEG and JPEG2000 lead to different locations in the same subband having an increase and decrease in entropy simultaneously. For example, in JPEG, increases in entropy occur due to the introduction of discontinuous blocking artifacts, while decreases in entropy occur in smoother regions that are heavily quantized. As a result, when weighted entropies are summed up, the gain and loss of entropies tend to cancel each other leading to lower quality ranges. When the entropies are summed up after the absolute value of the difference is calculated for every

block, the changes in the entropies are still preserved leading to an algorithm that possesses better overall performance.

As a limiting case of the above analysis, we analyze the performance of RRED indices that require only a single number from the reference. It turns out that these algorithms are more sensitive to the choice of the neural noise variance. Thus, we study the variation in performance of the single number algorithms as a function of the neural noise variance in Table 5.4.

Table 5.4: Effect of neural noise on RRED indices - SROCC

Distortion Type	$\sigma_W^2 = 0$	$\sigma_W^2 = 0.1$	$\sigma_W^2 = 1$
JPEG2000	0.9468	0.9514	0.9455
JPEG	0.9113	0.9152	0.9179
AWGN	0.9452	0.9447	0.9314
Gaussian Blur	0.9698	0.9038	0.6108
Fast fading errors	0.9181	0.9183	0.8773
Overall	0.7682	0.7978	0.8877

Observe that when $\sigma_W^2 = 0$, the RRED indices simply compute the difference between the scaled entropies of the reference and distorted wavelet coefficients. While the overall performance increases as σ_W^2 increases, the performance within the category of images distorted by Gaussian blur decreases. Thus it is possible to trade off these performances and achieve a desirable operating point by choosing the corresponding value of σ_W^2. This suggests that the neural noise variance of our model does impact the perception of blur apart from being able to better align scores belonging to different distortion categories. We could improve the overall performance for each value of σ_W^2 by sending one number per subband of the wavelet decomposition. For exam-

ple, by using 4 numbers from the four vertical subbands at different scales of the reference, we can improve the overall SROCC for $\sigma_W^2 = 0.1$ to 0.8606 by weighting the bands from coarser scales to the finer scales in the ratio 8:4:2:1. This algorithm still achieves excellent performance within each distortion category. Let RRED* denote this algorithm. We compare this algorithm against other popular RR QA algorithms in Table 5.5.

Table 5.5: Comparison of RRED* with other RR QA algorithms on LIVE Image Database - SROCC

Distortion Type	RRED*	Curvelet [24]	HWD2[24]	WNISM[93]
JPEG2000	0.9495	0.9170	0.9362	0.9135
JPEG	0.9512	0.9288	0.9543	0.9069
AWGN	0.9664	0.9585	0.9321	0.8703
Gaussian Blur	0.9453	0.9131	0.8282	0.9147
Fast fading errors	0.9310	0.9378	0.9386	0.9229
Overall	0.8606	0.9104	0.9418	0.7651
No. of scalars	4	24	16	18

We now present performance results for a couple of algorithms belonging to the framework of RRED indices on the Tampere Image Database 2008 (TID2008) [49]. TID2008 contains 17 types of distortions across 1700 distorted images. We use the same value of the neural noise $\sigma_W^2 = 0.1$ as used in all of the results presented in this section (other than the "single number" algorithms) on the LIVE Image Database. The results presented below demonstrate that the parameters of the algorithm do not require any training on databases. Table 5.6 contains results on the SROCC obtained over all on the database. MS-SSIM denotes the multiscale structural similarity index [92]. RRED$_{16}^{M_{16}}$

performs better than some of the best full reference algorithms while RRED_{22}^{1} is competitive with PSNR and outperforms [93].

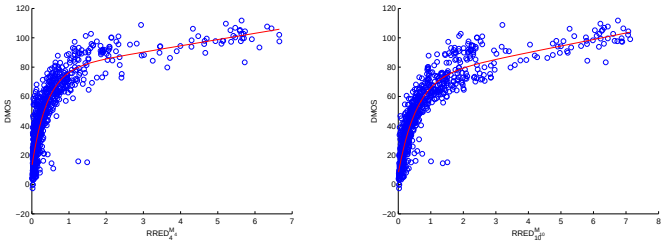
Table 5.6: Overall performance results on TID2008

Algorithm	SROCC
MS-SSIM	0.853
$\text{RRED}_{16}^{M_{16}}$	0.824
VIF	0.750
VSNR	0.705
PSNR	0.553
RRED_{22}^{1}	0.521
WNISM [93]	0.512

We conclude this section with a brief note on the computational complexity of the RRED algorithms. The steerable pyramid decomposition of the image has a computational complexity of $O(N \log N)$ arithmetic operations per scale, where N is the total number of pixels in the image. The following calculations represent the computational cost per scale. The estimation of $\mathbf{K}_U$ requires $O(N)$ operations, while the singular value decomposition of $\mathbf{K}_U$ requires $O(M^3)$ operations, where $\mathbf{K}_U$ is of size $M \times M$. Note that M represents the number of elements in a block of size $\sqrt{M} \times \sqrt{M}$. Further, the estimation of local variance parameters (premultipliers) requires $O(NM^2)$ operations, while sending the desired information from the reference or distorted image can be upper bounded by a cost of $O(N)$. Thus the overall computational complexity of the RRED algorithms is $O(N(\log N + M^2))$.

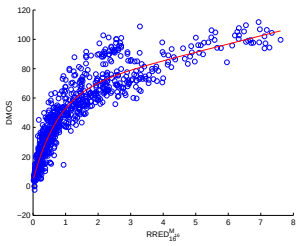
The time required for computing the index is calculated by running the algorithm on an Intel Pentium 4 processor with 2 GB RAM and 3.4 GHz

speed. The algorithm was simulated using MATLAB version R2008a without any optimization of the code. The total time required to compute the index can be divided into four different phases at both the reference and distorted image. The first phase involves reading the image which takes around 0.2 seconds. One of the bottle necks in the computation of the RRED indices is the time required to perform a multiscale multiorientation decomposition using steerable pyramids, which requires 1.5-1.6 seconds. Computing the scaled entropies for a subband, which is also the information required to be transmitted takes 0.9-1 seconds for the largest subband. This could be more depending on the number of subbands in which the index is evaluated. Finally, computation of the index from the numbers transmitted from the reference and distorted requires time of the order of milliseconds. Overall the algorithm requires around 2.8 seconds to transmit the desired information from each (reference or distorted) image.

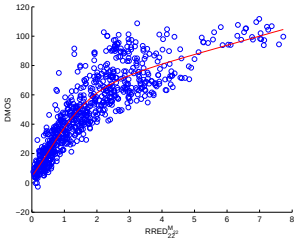


(a) $\text{RRED}_4^{M_4}$

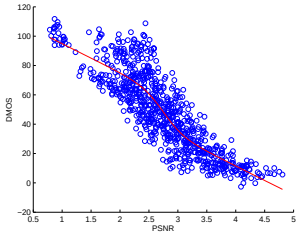
(b) $\text{RRED}_{10}^{M_{10}}$



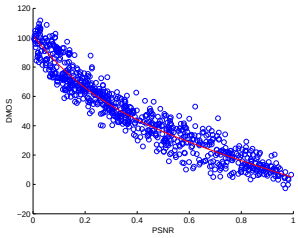
(c) $\text{RRED}_{16}^{M_{16}}$



(d) $\text{RRED}_{22}^{M_{22}}$



(e) PSNR



(f) VIF

Figure 5.2: Logistic fits used to compute linear correlation coefficient on the overall LIVE Image Quality Assessment Database for different objective algorithms

Chapter 6

Video Quality Assessment by Reduced Reference Spatio-temporal Entropic Differencing

6.1 Introduction

This chapter deals with reduced reference (RR) video quality assessment (QA), where only partial information from the reference can be made available in addition to the distorted video for quality evaluation. The problem of RR QA is particularly relevant in the context of video owing to the large amount of data involved. RR video quality assessment (VQA) is a challenging research problem for two reasons: the reduced reference constraint in the problem and the multidimensional (spatial and temporal) structure of the signal. Elaborating on the second aspect, VQA is significantly harder than IQA owing to the addition of a new dimension to the problem, namely the temporal dimension. A key attribute of the most successful VQA algorithms is their ability to capture *both* temporal and spatial distortions. Blocking artifacts, ringing effects and blur are examples of spatial distortions while jerkiness, ghosting and mosquito effects are examples of temporal distortions. [61] contains detailed descriptions of various spatial and temporal distortions that can afflict a video signal.

The spatial aspect of the problem has received a lot of attention over the years, and significant progress has been made owing to the availability of sophisticated models of the human visual system (HVS) and of natural scene statistics. However, the successful application of either HVS based approaches or statistical approaches to capture temporal distortions has been limited. While there do exist advanced models of motion perception, its applicability to VQA has not yet been completely realized. [90] and [61] represent examples of an effort towards this direction. On the other hand, statistical models of motion are almost non-existent in the literature and none has been found to be statistically regular over natural videos. Motion and other temporal changes in a video may be analyzed through optical flow vectors or by temporal filtering and decomposition. [54] provides a statistical model of optical flow vectors. However, the regularity of this model over all kinds of videos, including those that contain egomotion or significant movements over large areas of the frame appears to be difficult to verify. There is also little prior work on characterizing the statistics of temporally filtered natural videos. In this work, we build a statistical model of multiscale multiorientation wavelet decompositions of frame differences. This represents one of the first attempts at trying to obtain a statistical model of temporal changes in natural videos and to use them for VQA.

The main contributions of the chapter are as follows:

1. We present a natural video statistical model for the wavelet coefficients of frame differences between adjacent frames in a video sequence. These

coefficients possess a heavy tailed distribution and are well modeled by a Gaussian scale mixture (GSM) distribution.

2. The GSM model for the wavelet coefficients of frame differences are used to compute reduced reference entropic differences (RRED) between the reference and the distorted videos, leading to temporal RRED (or TRRED) indices. These indices are designed using a hybrid approach of statistical models and perceptual principles. The TRRED indices seek to measure the amount of motion information difference that occurs between the reference and distorted videos. We hypothesize that this information difference captures temporal distortions that can be perceived by humans.

3. The TRRED indices, in conjunction with our previously developed spatial RRED (or SRRED) indices evaluated by applying the RRED index in Chapter 5 on every frame of the video, yield the spatio-temporal RRED (STRRED) index, which performs very well on the LIVE Video Quality Assessment Database in terms of correlation with human judgments of quality.

4. A family of algorithms are developed that vary in the amount of information required from the reference for quality computation. In particular, 'single number algorithms' or those indices that require just a single number from the reference/distorted video per frame are developed that are shown to correlate well with human judgments of quality. The amount

of information can range up to almost full reference.

The SRRED, TRRED and STRRED indices designed here also possess all the other favorable properties that the RRED indices in Chapter 5 possess. In particular, they allow for bidirectional computation of quality, need not be trained on databases of human opinion scores and are applicable to scalable VQA. Depending on the desired application and problem constraints, this framework of algorithms also allows users to pick any algorithm from the class of algorithms presented that meets the performance requirements and data rates required for quality computation.

We now present a brief overview of prior work on video quality assessment. A detailed subjective evaluation of popular full reference video quality assessment algorithms can be found in [62]. Any image quality assessment algorithm can be trivially extended to videos by applying the algorithm on every frame of the video and calculating the average score. Thus successful image quality assessment (IQA) algorithms such as multiscale structural similarity index (SSIM) [92], Sarnoff just noticeable differences (JND) metrix [56] and visual signal to noise ratio (VSNR) [9] naturally lead to VQA algorithms. However, the performance of these algorithms is limited owing to their failure to capture temporal distortions in the video. Various researchers have developed QA algorithms that attempt to capture both spatial and temporal distortions. [90] computes a speed weighted SSIM index by associating weights to different locations depending on the amount of motion information.

The MOVIE index [61] uses the idea of motion tuning or the varied sensitivity of the humans to motion direction to measure distortions. Other FR VQA algorithms account for temporal distortions through temporal filters [95], temporal decorrelation [30], evaluating quality over temporal trajectories [6] and studying the temporal evolution of spatial distortions [43].

The Video Quality Metric (VQM) introduced by NTIA [48] is a reduced reference VQA algorithm that requires reference data of around 4% of the size of the uncompressed video sequence. The algorithm is based on computing losses in the spatial gradients of the luminance components and features based on the product of luminance contrast and motion, and by measuring color impairments. [39] is another RR VQA algorithm that calculates the weighted norm of the error in wavelet coefficients at selected locations to reduce the data required for quality computation. Other recent RR VQA algorithms include [28] and [53]. While the former is based on a discriminative analysis of harmonic strength, the latter is an extension of [90] to the reduced reference setting. The spatio-temporal RRED indices introduced in this dissertation offer significant improvements in performance and allow for excellent performance at low data rates of reference side information.

The rest of the chapter is organized as follows. We present a statistical model for the wavelet coefficients of frame differences in Section 6.2 followed by the system model on which the QA algorithms are based on in Section 6.3. We describe the spatio-temporal RRED indices in 6.4, provide a perceptual interpretation of the algorithms in 6.5 and discuss the performance of these

indices in Section 6.6.

6.2 Statistical Model of Frame Differences

We describe a statistical model of the wavelet coefficients (WC) of the frame differences between adjacent frames in a video sequence. Let all the wavelet coefficients of the frame differences in a subband be partitioned into non-overlapping blocks of size $\sqrt{N} \times \sqrt{N}$. Let the blocks in subband k, $k \in \{1, 2, \ldots, K\}$, be indexed by $m \in \{1, 2, \ldots, M_k\}$. Now the consider the wavelet coefficients of the frame differences between Frame f and $f + 1$ belonging to Block m of Subband k. Let $\bar{D}_{mkf}$ denote a vector of wavelet coefficients in Block m of Subband k and Frame f.

We model the block $\bar{D}_{mkf}$ as a Gaussian scale mixture (GSM) distributed vector with continuous scale. Specifically, $\bar{D}_{mkf}$ is distributed as

$$\bar{D}_{mkf} = T_{mkf}\bar{V}_{mkf}, \tag{6.1}$$

where T_{mkf} is independent of $\bar{V}_{kf}$ with $\bar{V}_{kf} \sim \mathcal{N}(0, \mathbf{K}_{V_{kf}})$. Note that we model $\bar{V}_{mkf}$ in every block of Subband k and Frame f to have the same covariance matrix $\mathbf{K}_{kf}$, with the premultiplier random variable T_{mkf} modulating the covariance matrix for different blocks. We also assume that T_{mkf} and $\bar{V}_{kf}$ are independent across all the indices m, k and f describing them.

Note that similar models describe the wavelet coefficients of frames in [84]. The GSM model implies that the divisively normalized coefficients (i.e. coefficients obtained by dividing $\bar{D}_{mkf}$ by T_{mkf}) follow a Gaussian distribu-

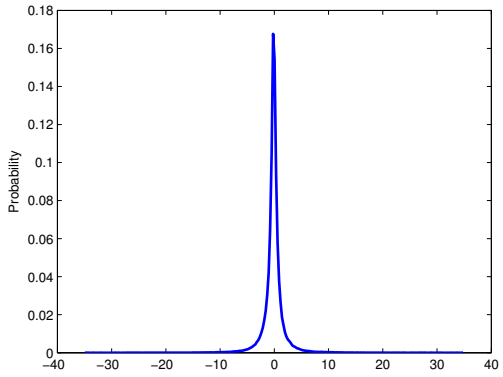
tion. This implication is clear since $\bar{D}_{mkf}/T_{mkf} = \bar{V}_{mkf}$, which is distributed as $\mathcal{N}(0, \mathbf{K}_{V_{kf}})$. We verify this empirically, by obtaining the maximum likelihood estimator of T_{mkf}, dividing the wavelet coefficient blocks by these and analyzing the distributions of normalized coefficients. The maximum likelihood estimator of T_{mkf} is given by [84],

$$\hat{T}_{mkf} = \underset{T_{mkf}}{\operatorname{argmax}}\, p(\bar{D}_{mkf}|T_{mkf}) = \sqrt{\frac{\bar{D}_{mkf}^T \mathbf{K}_{V_{kf}}^{-1} \bar{D}_{mkf}}{N}}.$$

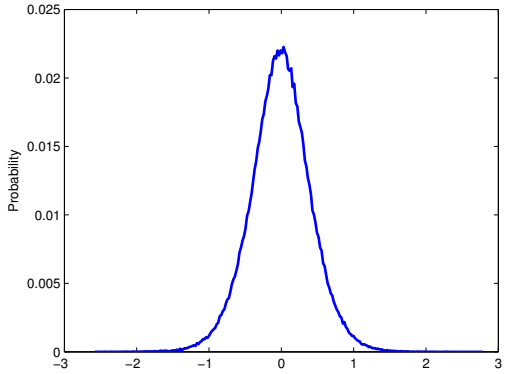
Table 6.1: Mean $\Delta H/H$ for reference videos on LIVE VQA database

Sequence	Mean $\Delta H/H$
1	0.0030
2	0.0019
3	0.0037
4	0.0024
5	0.0017
6	0.0025
7	0.0022
8	0.0024
9	0.0034
10	0.0026

In Fig. 6.1 we show that the empirical histograms of the WC of the frame differences are non-Gaussian and heavy tailed. In Fig. 6.2 we show that the corresponding normalized coefficients can be modeled well by a Gaussian distribution. The ratio of the relative entropy between the empirical distribution of the normalized coefficients and the Gaussian fitted distribution (denoted by ΔH) is calculated and shown to be a very small fraction of the entropy of the corresponding empirical distributions (denoted by H). Further,



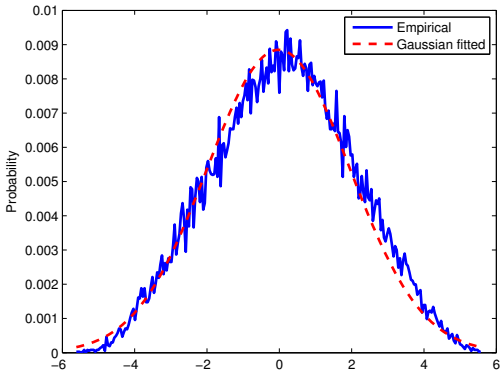
(a) Sunflower



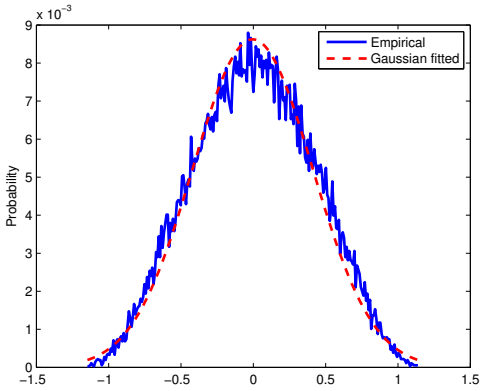
(b) Mobile Calendar

Figure 6.1: Distributions of the wavelet coefficients of the frame differences in a subband of reference videos.

in Table 6.1, we show that the ratio of relative entropy to the entropy of the empirical distribution ($\Delta H/H$) computed for each frame and averaged over all frames in every reference video sequence on the LIVE VQA database [62] is small. In the rest of the chapter, we refer to the local variances of the wavelet coefficients of the frame differences as 'local temporal variances'.



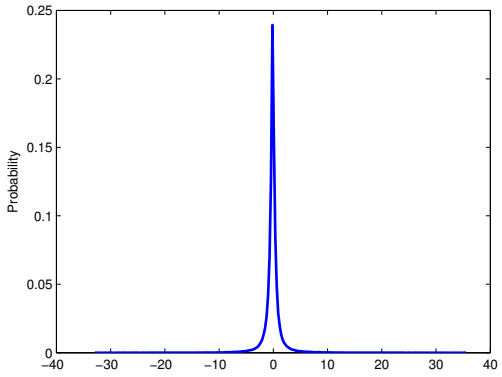
(a) Sunflower, $\frac{\Delta H}{H} = 0.0030$



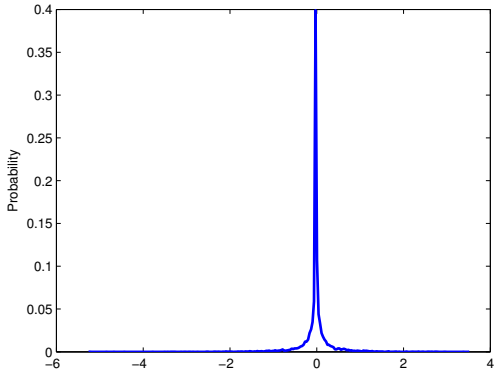
(b) Mobile Calendar, $\frac{\Delta H}{H} = 0.0029$

Figure 6.2: Empirical and Gaussian fitted statistics of divisively normalized wavelet coefficients (in a subband) of frame differences between 10th and 11th frame of 'sunflower' and 'mobile calendar' video sequences in LIVE Video Quality Assessment Database. ΔH denotes the relative entropy between the empirical distribution and the Gaussian fitted distribution while H denotes the entropy of the empirical distribution of the normalized coefficients.

In Fig. 6.3 and Fig. 6.4, we show exemplar corresponding plots (of the empirical histograms and the normalized coefficients) for videos distorted by H.264 compression. It is evident that the normalized coefficients of 'sunflower'



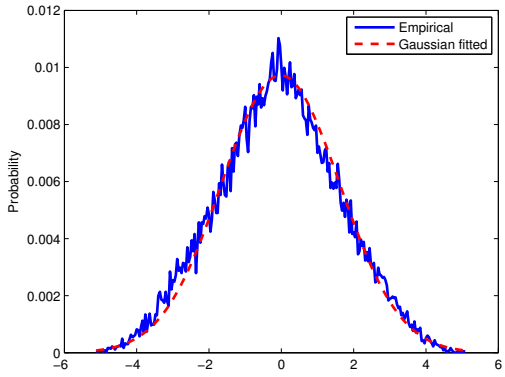
(a) Sunflower



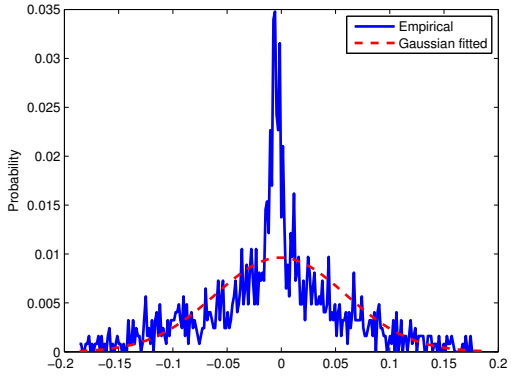
(b) Mobile Calendar

Figure 6.3: Distributions of the wavelet coefficients of the frame differences in a subband of H.264 compressed videos.

appear to be Gaussian distributed while those of 'mobile calendar' are not Gaussian distributed. While we do not explicitly exploit the Gaussianity or non-Gaussianity of the normalized coefficients of the distorted videos in our algorithms, we use a GSM model for the distorted video's temporal coefficients as was done in Chapter 5 for the spatial coefficients of distorted images.



(a) H.264 compressed Sunflower



(b) H.264 compressed Mobile Calendar

Figure 6.4: Empirical and Gaussian fitted statistics of divisively normalized wavelet coefficients (in a subband) of frame differences between 10th and 11th frame of H.264 compressed 'sunflower' and 'mobile calendar' video sequences in LIVE Video Quality Assessment Database.

By fitting a GSM model for the distorted video, we compute the natural approximation of the distorted video. If the normalized coefficients are indeed Gaussian distributed, then it means that the natural approximation is close to the empirical distributions of the distorted coefficients. The entropy difference

between the reference and the distorted videos can then be interpreted as a distance between the reference video and the natural approximation of the distorted video. Our model is based on the hypothesis that such a distance is meaningful for perceptual quality and this is validated by the good performance of our algorithms on the LIVE VQA database. This also avoids the need for learning distortion-specific statistical models.

6.3 System Model

Let $\bar{C}_{mkfr}$ and $\bar{C}_{mkfd}$ denote a vector of wavelet coefficients in Block m, Subband k and Frame f of the reference and distorted video respectively. On account of the GSM model for wavelet coefficients, we have

$$\bar{C}_{mkfr} = S_{mkfr}\bar{U}_{mkfr} \qquad \bar{C}_{mkfd} = S_{mkfd}\bar{U}_{mkfd},$$

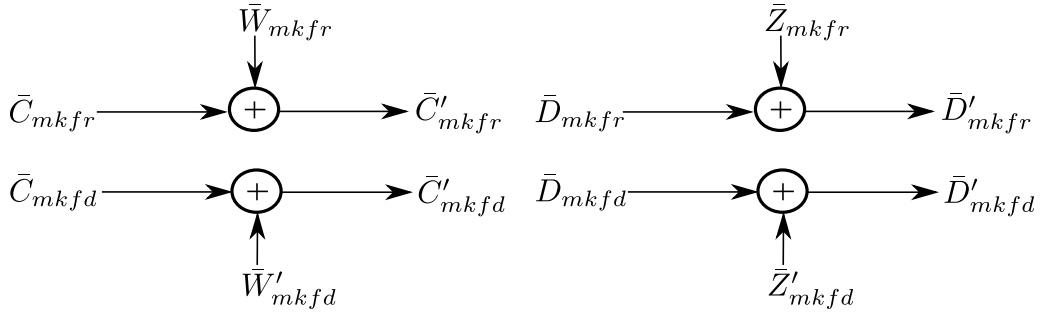
where S_{mkfr} is independent of $\bar{U}_{mkfr}$, S_{mkfd} is independent of $\bar{U}_{mkfd}$, $\bar{U}_{mkfr} \sim \mathcal{N}(0.\mathbf{K}_{U_{kfr}})$, $\bar{U}_{mkfd} \sim \mathcal{N}(0.\mathbf{K}_{U_{kfd}})$ and S_{mkfr} and S_{mkfd} are non-negative random variables. We assume that all the blocks are independent of each other in order to simplify the index. Inter-block interactions are not accounted for in this index, although it is likely possible to obtain better indices by exploiting such interactions as well given that accurate enough models could be found to apply. Note that as in Chapter 5, we use a natural image approximation for the WC in each frame of the distorted video as well and measure quality using this natural video approximation.

Let $\bar{D}_{mkfd}$ and $\bar{D}_{mkfd}$ be a vector of wavelet coefficients of frame differ-

ences between Frame f and Frame $f+1$ in Block m and Subband k belonging to the reference and distorted video respectively. Using the model in (6.1) for the WC of frame differences, we have

$$\bar{D}_{mkfr} = T_{mkfr}\bar{V}_{mkfr} \qquad \bar{D}_{mkfd} = T_{mkfd}\bar{V}_{mkfd},$$

where T_{mkfr} is independent of $\bar{V}_{kfr}$, T_{mkfd} is independent of $\bar{V}_{kfd}$, $\bar{V}_{kfr} \sim \mathcal{N}(0.\mathbf{K}_{V_{kfr}})$, $\bar{V}_{kfd} \sim \mathcal{N}(0.\mathbf{K}_{V_{kfd}})$ and T_{mkfr} and T_{mkfd} are non-negative random variables. Again as before, all the blocks are assumed independent of each other.



(a) Wavelet coefficients of frames (b) Wavelet coefficients of frame differences

Figure 6.5: System model for computation of SRRED and TRRED indices

We allow the wavelet coefficients in each block of the reference and distorted video frames as well as the wavelet coefficients of the frame differences corresponding to each block of the reference and distorted frames to pass through a Gaussian channel in order to model imperfections in the visual perception of these coefficients in the human visual system, eg. neural noise [63]. These models are shown in Fig. 6.5 and are expressed as

$$\bar{C}'_{mkfr} = \bar{C}_{mkfr} + \bar{W}_{mkfr} \qquad \bar{C}'_{mkfd} = \bar{C}_{mkfd} + \bar{W}_{mkfd},$$

where $\bar{W}_{mkfr} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$ and $\bar{W}_{mkfd} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_N)$.

Similarly, we have

$$\bar{D}'_{mkfr} = \bar{D}_{mkfr} + \bar{Z}_{mkfr} \qquad \bar{D}'_{mkfd} = \bar{D}_{mkfd} + \bar{Z}_{mkfd},$$

where $\bar{Z}_{mkfr} \sim \mathcal{N}(0, \sigma_Z^2 \mathbf{I}_N)$ and $\bar{Z}_{mkfd} \sim \mathcal{N}(0, \sigma_Z^2 \mathbf{I}_N)$.

The RRED indices essentially compute the differences of scaled entropies of the neural noisy wavelet coefficients of frames or neural noisy wavelet coefficients of the frame differences. These are described in detail in the following section.

6.4 Spatio-temporal RRED Indices

We first describe the design of the spatial RRED indices and temporal RRED indices and then show how these may be combined to yield spatio-temporal RRED indices.

6.4.1 Spatial RRED (SRRED)

The spatial RRED or SRRED indices are computed in a manner similar to Chapter 5. We evaluate the entropies of $\bar{C}'_{mkfr}$ and $\bar{C}'_{mkfd}$ conditioned on the maximum likelihood estimates of S_{mkfr} and S_{mkfd} respectively. Let s_{mkfr} and s_{mkfd} be the maximum likelihood estimates of S_{mkfr} and S_{mkfd} given the corresponding frames in the reference and distorted video, respectively. The entropies of $\bar{C}'_{mkfr}$ and $\bar{C}'_{mkfd}$ conditioned on $S_{mkfr} = s_{mkfr}$ and $S_{mkfd} = s_{mkfd}$

are given by

$$h(\bar{C}'_{mkfr}|S_{mkfr} = s_{mkfr}) = \frac{1}{2}\log\left[(2\pi e)^N|s^2_{mkfr}\mathbf{K}_{U_{kfr}} + \sigma^2_W\mathbf{I}_N|\right]$$
$$h(\bar{C}'_{mkfd}|S_{mkfd} = s_{mkfd}) = \frac{1}{2}\log\left[(2\pi e)^N|s^2_{mkfd}\mathbf{K}_{U_{kfd}} + \sigma^2_W\mathbf{I}_N|\right].$$

Let the scaling factors be defined by

$$\gamma_{mkfr} = \log(1 + s^2_{mkfr}) \qquad \gamma_{mkfd} = \log(1 + s^2_{mkfd}). \tag{6.2}$$

The scalars defined in (6.2) are exactly the same as used in Chapter 5 and enhance the local nature of the algorithm, allow for variable weighting of the amount of visual information in different regions of each frame and embed numerical stability in the algorithm for small values of the neural noise variance and the local spatial variance estimate.

The SRRED index in Subband k obtained by using M_k scalars (one for each block) from the reference and distorted is given by

$$\text{SRRED}_k^{M_k} = \frac{1}{FM_k}\sum_{f=1}^{F}\sum_{m=1}^{M_k}\Big|\gamma_{mkfr}h(\bar{C}'_{mkfr}|S_{mkfr} = s_{mkfr}) - \gamma_{mkfd}h(\bar{C}'_{mkfd}|S_{mkfr} = s_{mkfd})\Big|.$$

The number of scalars required in Subband k can be reduced by summing the entropy terms over small patches and sending these partial sums. The number of scalars required is equal to the number of patches. As an extreme case, all of the entropy terms can be summed up and sent as a single number from the reference. The SRRED index in Subband k using 1 scalar is given by

$$\text{SRRED}_k^1 = \frac{1}{FM_k}\sum_{f=1}^{F}\Big|\sum_{m=1}^{M_k}\gamma_{mkfr}h(\bar{C}'_{mkfr}|S_{mkfr}=s_{mkfr}) - \sum_{m=1}^{M_k}\gamma_{mkfd}h(\bar{C}'_{mkfd}|S_{mkfr}=s_{mkfd})\Big|.$$

6.4.2 Temporal RRED or TRRED

The temporal RRED or TRRED indices are obtained by computing the differences of scaled conditional entropies of the wavelet coefficient differences. Denote t_{mkfr} and t_{mkfd} as the maximum likelihood estimates of T_{mkfr} and T_{mkfd} given the corresponding adjacent frames in the reference and distorted video respectively. The entropies of $\bar{D}'_{mkfr}$ and $\bar{D}'_{mkfd}$ conditioned on $T_{mkfr} = t_{mkfr}$ and $T_{mkfd} = t_{mkfd}$ are given by

$$h(\bar{D}'_{mkfr}|T_{mkfr}=t_{mkfr}) = \frac{1}{2}\log\left[(2\pi e)^N|t_{mkfr}^2\mathbf{K}_{V_{kfr}} + \sigma_Z^2\mathbf{I}_N|\right]$$
$$h(\bar{D}'_{mkfd}|T_{mkfd}=t_{mkfd}) = \frac{1}{2}\log\left[(2\pi e)^N|t_{mkfd}^2\mathbf{K}_{V_{kfd}} + \sigma_Z^2\mathbf{I}_N|\right].$$

Now define the scaling factors

$$\delta_{mkfr} = \log(1+s_{mkfr}^2)\log(1+t_{mkfr}^2) \qquad \delta_{mkfd} = \log(1+s_{mkfd}^2)\log(1+t_{mkfd}^2), \tag{6.3}$$

which attach different importances to the amount of visual information at different locations and depend on both the local temporal as well as spatial variance parameters in contrast to the scaling factors defined in (6.2) which only depend on the spatial variance parameters. The scalars used in (6.3) are

a product of two components. The first component, which is a function of the local variance of the spatial decomposition of the frame can be interpreted as follows. [75, 76] suggest that at slow speeds, lowering the contrast lowers the perceived speed. Under the assumption that the speeds due to frames sampled at practical temporal sampling rates (between 25 frames per second and 50 frames per second) can be considered slow speeds, we believe that the factor $\log(1 + s^2_{mkfr})$ (or $\log(1 + s^2_{mkfd})$), which is an increasing function of the local spatial variance, lowers the perceived speed as the contrast reduces. The local spatial variance s^2_{mkfr} (or s^2_{mkfd}) measures the spatial luminance contrast in different regions of the frame. The other component of the scaling factor, $\log(1 + t^2_{mkfr})$ (or $\log(1 + t^2_{mkfd})$), which depends on the local temporal variance, has an effect similar to the effect of local spatial variances for SRRED indices, where it embeds local nature to the algorithm and allows weighting of the temporal information from different locations according to the amount of local temporal variance.

The TRRED index in Subband k obtained by using M_k scalars from the reference and distorted video difference frames is given by

$$\mathrm{TRRED}_k^{M_k} = \frac{1}{FM_k} \sum_{f=1}^{F} \sum_{m=1}^{M_k} \Big| \delta_{mkfr} h(\bar{D}'_{mkfr} | T_{mkfr} = t_{mkfr}) - \delta_{mkfd} h(\bar{D}'_{mkfd} | T_{mkfr} = t_{mkfd}) \Big|.$$

Similar to the SRRED indices, the amount of information required in Subband k of Frame f for the TRRED index can be reduced by summing the entropy terms over small patches. In the limiting case, we obtain 'single

number algorithms', where we require one number per frame for evaluation of the TRRED index. This may be represented as

$$\mathrm{TRRED}_k^1 = \frac{1}{FM_k}\sum_{f=1}^{F}\Big|\sum_{m=1}^{M_k}\delta_{mkfr}h(\bar{D}'_{mkfr}|T_{mkfr}=t_{mkfr}) - \sum_{m=1}^{M_k}\delta_{mkfd}h(\bar{D}'_{mkfd}|T_{mkfr}=t_{mkfd})\Big|.$$

6.4.3 Spatio-temporal RRED or STRRED

The spatio-temporal RRED indices combine the SRRED and the TRRED indices. Note that the SRRED and TRRED indices operate individually on data obtained by separate processing of the spatial and temporal frequency components. This matches well-accepted models of separable spatial and temporal frequency tuning of area V1 neurons [29, 68]. According to this model, the response of area V1 neurons to temporal frequencies is not affected by the spatial frequency of the stimulus and vice versa. A product form can thus be used to represent separable spatial and temporal frequency responses of the area V1 cortical neurons. We model the separable processing of area V1 neurons as opposed to [61], which models the behavior of the neurons in area middle temporal (MT). The area MT neurons are tuned to velocity (both direction and speed) and their selectivity towards motion direction and speed inhibits the spatio-temporal frequency separability. By matching the separable part of the cortical processing, we are able to capture 'pure temporal' distortions which are often 'flickery' without computing motion vectors. This is perhaps particularly relevant and important in the context of QA.

Interestingly, while the SRRED indices are obtained using only spatial frequency information, the TRRED indices are obtained using spatial and temporal information (the spatial information is used to weight the temporal information). As a result, only the TRRED indices are influenced by temporal distortions while both SRRED and TRRED indices are affected by spatial distortions. The computation of the quality index from the spatial and temporal information concerns the processing that occurs in the later stages of human visual processing, where there is evidence of interactions between the two [68]. While we are inspired by these observations, and have used them in constructing our RR VQA models, we do not claim to replicate specific cortical processing modules.

The spatio-temporal RRED index is obtained as a product of the spatial and temporal RRED indices and is expressed as

$$\text{STRRED}_k = \text{SRRED}_k \text{TRRED}_k.$$

6.4.4 Estimation of Parameters

The maximum likelihood (ML) estimates of the local spatial and temporal variances as well as the covariance matrices in a given subband are given below. Derivations of these estimates can be found in [84] and [63]. The estimates of the spatial and temporal covariance matrices of Subband k in Frame

f of the reference and distorted videos are given by

$$\hat{\mathbf{K}}_{U_{kfr}} = \sum_{m=1}^{M_k} \frac{\bar{C}_{mkfr}\bar{C}_{mkfr}^T}{M_k} \qquad \hat{\mathbf{K}}_{V_{kfr}} = \sum_{m=1}^{M_k} \frac{\bar{D}_{mkfr}\bar{D}_{mkfr}^T}{M_k}$$
$$\hat{\mathbf{K}}_{U_{kfd}} = \sum_{m=1}^{M_k} \frac{\bar{C}_{mkfd}\bar{C}_{mkfd}^T}{M_k} \qquad \hat{\mathbf{K}}_{V_{kfd}} = \sum_{m=1}^{M_k} \frac{\bar{D}_{mkfd}\bar{D}_{mkfd}^T}{M_k}.$$

Similarly, the ML estimates of the local spatial and temporal variances of the reference and distorted videos are given by

$$\hat{s}_{mkfr}^2 = \frac{\bar{C}_{mkfr}^T\mathbf{K}_{U_{kfr}}^{-1}\bar{C}_{mkfr}}{N} \qquad \hat{t}_{mkfr}^2 = \frac{\bar{D}_{mkfr}^T\mathbf{K}_{V_{kfr}}^{-1}\bar{D}_{mkfr}}{N}$$
$$\hat{s}_{mkfd}^2 = \frac{\bar{C}_{mkfd}^T\mathbf{K}_{U_{kfd}}^{-1}\bar{C}_{mkfd}}{N} \qquad \hat{t}_{mkfd}^2 = \frac{\bar{D}_{mkfd}^T\mathbf{K}_{V_{kfd}}^{-1}\bar{D}_{mkfd}}{N}.$$

6.5 Perceptual Properties of Spatio-Temporal RRED Indices

We now discuss perceptual properties that the spatio-temporal RRED indices rely upon. See the previous chapter for more discussion on perceptual aspects of the algorithm that relate to the STRRED indices developed here.

6.5.1 Spatial and Temporal Multiscale Multiorientation Decomposition

The spatio-temporal RRED indices involve separate spatial and temporal decompositions of the given video motivated by the evidence of mostly separable processing of the spatial and temporal data in the visual cortex. In particular, the SRRED indices are computed using spatial multiscale multiorientation decompositions of the each frame in the video sequence, while the

TRRED indices are computed using a spatial multiscale multiorientation decomposition of the frame differences. Frame differences capture the temporal information in the video and we further subject it to a multiscale multiorientation decomposition before computing the index. We discuss the perceptual implications of each of these aspects in the following.

The SRRED indices are computed according to Chapter 5 and therefore involve a multiscale multiorientation decomposition of the given image before the index is computed. There is ample evidence in the visual science literature that suggests similar signal processing occurs in the early stages of visual processing. We refer the reader to [61, 70] for a detailed account of such signal processing and a comparison of various filters used in this process. Here, we simply mention that steerable pyramids are used for the multiscale multiorientation spatial wavelet decomposition as opposed to the spatio-temporal Gabor filters used in [61].

We capture temporal information present in the video signal through frame differences and subject them to a multiscale multiorientation wavelet decomposition using steerable pyramids. The wavelet coefficients of frame differences may be interpreted as 'moving edges' and we substantiate this conclusion through the experimental results shown in Table 6.2. The objective of the experiment is to find out the relationship between the local variances of the wavelet coefficients of the frame differences (denoted by T and referred to as local temporal variances), the local variances of wavelet coefficients of frames (denoted by S and referred to as local spatial variances) and the optical flow

evaluated at each block corresponding to S and T using the algorithm in [73] (denoted by OF). Note that $\mathbb{E}[T] \sim 1$ as indicated in the estimation procedure in the previous section. We show through Table 6.2 that $\mathbb{E}[S|T > 1] \geq \mathbb{E}[S]$ and $\mathbb{E}[OF|T > 1] \geq \mathbb{E}[OF]$ for the 10 diverse reference videos on the LIVE Video Quality Assessment Database. In the table, empirical estimates of $\mathbb{E}[S]$ and $\mathbb{E}[S|T > 1]$ are obtained for each frame after computing the ML estimate of S and T at every block, and the estimates of the expectations in each frame are again averaged across frames for every reference video. The flow estimates are obtained using the optical flow algorithm [73] and the empirical expectations are evaluated in a similar manner. Similar behavior is obtained even when the expectations are conditioned on the event $T > c$ with $c > 1$. These results suggest that given that T is large, S and OF are also large. S is large at high frequency locations corresponding to edges, while OF is large in locations where there is motion. Thus T being high implies that S and OF are high. Therefore locations where T is high may be interpreted as 'moving edges'.

6.5.2 Effect of Motion on Spatio-temporal Information

Motion tends to have the effect of shifting the frequency response curves down the scale of the wavelet decomposition. According to [7], "*motion does not diminish the visual passband, but instead slides the spatial frequency window down the spatial frequency scale.*" In other words, in the presence of motion, the humans are more sensitive to the wavelet decomposition coefficients

at the coarser scales than at finer scales. We enforce this perceptual phenomenon in our spatio-temporal RRED indices by computing the SRRED and the TRRED indices in the coarsest scale of the steerable pyramid decomposition of the frames and the frame differences respectively. We also observed empirically, that indices computed at these scales gave the best performance.

6.6 Results and Discussion

We now present results of the correlation analysis of the spatio-temporal RRED indices on the LIVE Video Quality Assessment Database. The LIVE Video Quality Assessment Database contains 10 reference videos and 150 distorted videos spanning 4 categories of distortions, including compression artifacts due to MPEG and H.264, errors induced by transmission over IP networks and errors introduced due to transmission over wireless networks. 6 videos contain 250 frames at 25 fps, 1 video contains 217 frames at 25 fps and

Table 6.2: Relation between T, S and OF

Sequence	$\mathbb{E}[S]$	$\mathbb{E}[S\|T>1]$	$\mathbb{E}[OF]$	$\mathbb{E}[OF\|T>1]$
1	0.97	3.38	2.01	3.61
2	0.99	2.60	2.98	4.95
3	1.01	4.49	0.78	1.64
4	1.00	2.01	4.03	4.05
5	1.01	2.82	0.83	1.15
6	1.02	3.85	2.58	2.77
7	1.02	2.67	2.21	3.14
8	1.01	2.06	1.33	1.41
9	1.01	2.80	0.97	1.10
10	1.01	2.16	1.04	1.06

3 videos contain 500 frames at 50 fps.

6.6.1 Implementation Details

The luminance frames in the video sequence, as well as the luminance frame differences, are subjected to a multiscale multiorientation wavelet decomposition using steerable pyramids [67]. The decomposition is performed at 3 scales and 6 orientations. Every subband is partitioned into non-overlapping blocks, each of size 3×3. The value of the neural noise variance for both the spatial and temporal data is chosen to be 0.1, i.e. $\sigma_W^2 = \sigma_Z^2 = 0.1$. Note that similar values of the neural noise variance were chosen in [70] and [63].

6.6.2 Subjective Evaluation

Table 6.3: SROCC between spatio-temporal RRED indices at different orientations in the coarsest scale and LIVE Video Quality Assessment Database scores

Index	SROCC
$\text{STRRED}_2^{M_2}$	0.7508
$\text{STRRED}_3^{M_3}$	0.7958
$\text{STRRED}_4^{M_4}$	0.8007
$\text{STRRED}_5^{M_5}$	0.7674
$\text{STRRED}_6^{M_6}$	0.7330
$\text{STRRED}_7^{M_7}$	0.7187

The SRRED, TRRED and STRRED indices are evaluated against subjective quality scores on the LIVE Video Quality Assessment Database. As mentioned earlier, the wavelet coefficients in the coarsest passband (for both the decomposition of the frames as well as the decomposition of the frame dif-

ferences) yield the best performance and all the performance results reported in this section are based on these wavelet coefficients. Further, we compare the performance of the spatio-temporal RRED indices evaluated in different orientations at the coarsest scale in Table 6.3. We use the Spearman Rank Order Correlation Coefficient (SROCC) between the subjective scores and the quality indices to compare the relative performances between the orientations. $\text{STRRED}_4^{M_4}$ yields the best performance among different orientations and this corresponds to the vertically oriented subband. In the rest of this section, we present detailed comparisons of the SRRED, TRRED and STRRED indices evaluated in the vertically oriented subband at the coarsest scale against other VQA algorithms.

Table 6.4 contains a detailed comparison of the SROCC of those RRED indices that operate at a high data rate against PSNR, multiscale (MS)-SSIM (computed for every frame and averaged across frames) [92], VQM [48] and the MOVIE index [61] against human opinion scores available with the LIVE VQA database [62].

The results in Table 6.4 show that the STRRED algorithms using $L/576$ scalars perform as well as some of the best FR VQA algorithms such as the MOVIE index. A similar behavior is also observed for the SROCCs of the single number algorithms presented in Table 6.5. We recall that these single number algorithms require just a single number per frame from the reference/distorted for quality computation. It is clear from the results, that even the single number algorithms significantly outperform PSNR, which is in fact a

Table 6.4: SROCC between spatio-temporal RRED indices, PSNR, MS-SSIM, VQM, MOVIE and LIVE Video Quality Assessment Database scores. NOSPF - Number of scalars per frame

Distortion Type	Wireless	IP	H.264	MPEG	Overall	NOSPF
$\text{STRRED}_4^{M_4}$	0.7857	0.7722	0.8193	0.7193	0.8007	L/576
$\text{SRRED}_4^{M_4}$	0.7925	0.7624	0.7542	0.7249	0.7592	L/1152
$\text{TRRED}_4^{M_4}$	0.7765	0.7513	0.8189	0.5879	0.7802	L/1152
PSNR	0.6574	0.4167	0.4585	0.3862	0.5398	L
MS-SSIM	0.7289	0.6534	0.7313	0.6684	0.7364	L
VQM	0.7214	0.6383	0.6520	0.7810	0.7026	L/25
MOVIE	0.8114	0.7192	0.7797	0.8170	0.8055	L

Table 6.5: SROCC between single number spatio-temporal RRED indices and LIVE Video Quality Assessment Database scores. NOSPF - Number of scalars per frame

Distortion Type	Wireless	IP	H.264	MPEG	Overall	NOSPF
STRRED_4^1	0.7208	0.5075	0.7197	0.7247	0.7319	1
$\text{SRRED}_4^{1/2}$	0.6066	0.3851	0.4441	0.7540	0.5961	1/2
$\text{TRRED}_4^{1/2}$	0.5863	0.5279	0.6737	0.4363	0.5870	1/2

FR quality index. The performance of the single number algorithms is almost on par with some of the FR VQA algorithms such as MS-SSIM. It is interesting that while the individual performances of the single number versions of the SRRED and TRRED indices are not outstanding, it is their combination that renders the algorithm effective. It appears that SRRED and TRRED seem to be capturing complementary distortions and that their combination makes them highly competitive.

Table 6.6: LCC between spatio-temporal RRED indices, PSNR, MS-SSIM, VQM, MOVIE and LIVE Video Quality Assessment Database scores. NOSPF - Number of scalars per frame

Distortion Type	Wireless	IP	H.264	MPEG	Overall	NOSPF
$\text{STRRED}_4^{M_4}$	0.8039	0.8020	0.8228	0.7467	0.8062	L/576
$\text{SRRED}_4^{M_4}$	0.8067	0.8033	0.7462	0.7281	0.7764	L/1152
$\text{TRRED}_4^{M_4}$	0.7726	0.7619	0.8324	0.5998	0.7743	L/1152
PSNR	0.6695	0.4689	0.5330	0.3986	0.5604	L
MS-SSIM	0.7157	0.7267	0.7020	0.6640	0.7379	L
VQM	0.7325	0.6480	0.6459	0.7860	0.7326	L/25
MOVIE	0.8371	0.7383	0.7920	0.8252	0.8217	L

Table 6.7: LCC between single number spatio-temporal RRED indices and LIVE Video Quality Assessment Database scores. NOSPF - Number of scalars per frame

Distortion Type	Wireless	IP	H.264	MPEG	Overall	NOSPF
STRRED_4^{1}	0.7051	0.5453	0.7242	0.7490	0.7264	1
$\text{SRRED}_4^{1/2}$	0.6218	0.4148	0.4270	0.7605	0.6057	1/2
$\text{TRRED}_4^{1/2}$	0.5991	0.5096	0.7058	0.4292	0.5741	1/2

We observe that for MPEG distortions, the single number RRED indices appear to perform better than the ones that use more information from

the reference. This occurs because the pooling strategies of the single number algorithms and the ones that use more reference information are different. In principle, the pooling strategy employed by the single number algorithms can also be used by the algorithms that operate using more information from the reference. However, their overall performance would be poorer if such a pooling strategy were to be performed. We would like to clarify that while the algorithms that use more information from the reference can use the pooling strategy of the single number algorithms, the reverse is not possible. The single number algorithms do not have enough information to employ any other strategy, since they are supplied with just one scalar.

We report results of linear correlation coefficient (LCC) scores in Tables 6.6 and 6.7. We use a nonlinearity on the objective scores before computing the LCC. For the spatio-temporal RRED indices, we use the nonlinearity given by

$$\text{Quality}(x) = \beta_1 \log(1 + \beta_2 x).$$

The fit between the subjective and objective scores used to compute the LCC for the spatio-temporal RRED indices is shown in Fig. 6.6. The LCC scores also follow similar trends as compared with the SROCC scores.

While in Table 6.4 and Table 6.5 we discuss the performance of spatio-temporal RRED indices that operate at two extremes of the amount of information from the reference, in Table 6.8, we show the variation in the performance as we increase the amount of information. The SROCC computed on the overall database is shown for the spatio-temporal RRED indices. The

amount of information is reduced by computing and sending partial sums of the scaled local entropies instead of the scaled entropies at all the locations as in [70]. The table shows that the algorithm which requires $L/27648$ scalars from the reference achieves a performance that is almost as good as the algorithm that requires $L/576$. We denote $L' = L/576$ where L is the number of pixels in a frame of the video. Note that $0.002L'$ corresponds to the 'single number' algorithm which requires just a single scalar from the reference per frame. The minor increase in the performance of the STRRED indices with decrease of information between rows 1 and 2 in Table 6.8 is due to the difference between computing the sum of the absolute differences of partial sums and the sum of the absolute differences of the scaled entropies at different locations. This phenomenon was also explained earlier in this section with regards to the superior performance of the single number algorithms for MPEG distortions over the algorithms that operate at higher data rates of reference information.

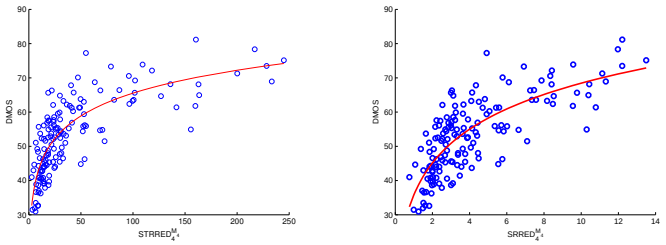
Table 6.8: Variation of overall SROCC with the amount of information required from the reference for the spatio-temporal RRED indices

No. of scalars per frame	SROCC
L'	0.8007
$0.250\ L'$	0.8056
$0.167\ L'$	0.8056
$0.083\ L'$	0.8015
$0.042\ L'$	0.7992
$0.021\ L'$	0.7930
$0.014\ L'$	0.7838
$0.010\ L'$	0.7660
$0.002\ L'$	0.7319

6.6.3 Computational Complexity and Run Times

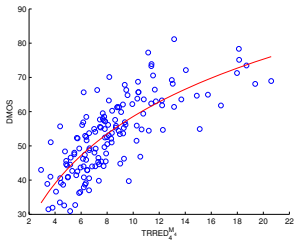
The computational complexity of the spatio-temporal RRED indices closely follows that of the RRED indices Chapter 5. Both the SRRED and the TRRED indices require $O(N(\log N + M^2))$ operations per frame per subband, where $\sqrt{M} \times \sqrt{M}$ is the size of a block during the processing of each subband. We refer the reader to Chapter 5 for a detailed calculation. Technically, the TRRED indices require a differencing step prior to computation of the wavelet transform, which requires N more operations and is consumed in the order notation. Thus, the overall complexity of the spatio-temporal RRED indices may be written as $O(FN(\log N + M^2))$, where F is the number of frames.

We also report run time results of experiments conducted on an Intel Pentium processor with 2 GB RAM and 3.4 GHz speed using MATLAC (R2008a) without code optimization. For a sequence of length 250 frames, the multiscale multiorientation transform using steerable pyramids takes around 215 seconds while the computation of the scaled entropy information which is finally differenced takes 4-6 seconds for any (reference or distorted) video. Note that the steerable pyramid decomposition time reported above is the time taken to compute the decomposition into 26 subbands (at 4 scales and 6 orientations). By operating the spatio-temporal RRED algorithms in only one of the subbands, it is possible to reduce the time required for this step. Despite this, it appears that the steerable pyramid decomposition is the bottleneck and improving the efficiency of this step will help improve the overall run time of the algorithm.

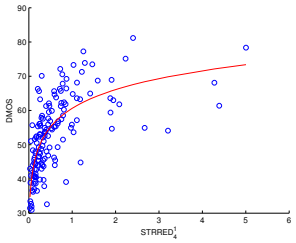


(a) STRRED$_4^{M_4}$

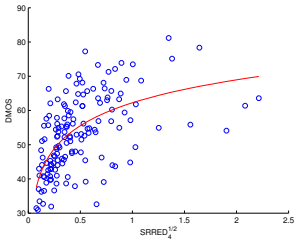
(b) SRRED$_4^{M_4}$



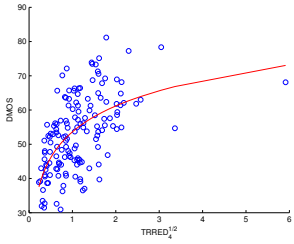
(c) TRRED$_4^{M_4}$



(d) STRRED$_4^1$



(e) SRRED$_4^{1/2}$



(f) TRRED$_4^{1/2}$

Figure 6.6: Nonlinear fit used to compute linear correlation coefficient on the overall LIVE Video Quality Assessment Database

Chapter 7

Conclusion and Future Work

We considered sophisticated problems in distributed compression and quality assessment, designed novel algorithms and demonstrated their efficiency through detailed analyses. In the domain of distributed compression, we took important steps towards better understanding the fundamental trade-offs in arbitrary networks by studying new non-trivial multiterminal settings. We also addressed the challenging problem of reduced reference quality assessment by developing a framework of algorithms and thereby bringing to light the trade-off between the amount of reference information and performance in terms of correlation with human judgments of quality.

We introduced the vacationing-CEO problem which in essence, is a CEO problem with multiple descriptions. We described a Gaussian achievable scheme and presented two lower bounds for the sum rate as optimization problems over the code parameters. We also showed that the Gaussian scheme is optimal in terms of sum rate for a class of distortion constraints, while time sharing is optimal for another class of distortion constraints. Since the time sharing scheme involves Gaussian schemes that may not individually satisfy the distortion constraints, a clear characterization of all such rates could help

prove optimality in general. We might also need better achievable schemes. Future work includes extending the achievable schemes to a more general two terminal source coding problem with multiple descriptions where the observations at both terminals are required to be recovered within their own distortion criteria at each receiver.

We analyzed a quadratic Gaussian multi-terminal source coding problem where two sources are to be compressed and delivered to the decoder through a relay. We presented achievable strategies based on compute-and-forward and compress-and-forward principles using lattices. Depending on problem parameters, either of these strategies could outperform the other in terms of the sum rate of the system. For the symmetric case, we found that the lower bound and achievable sum rate we obtain are within 1/2 bits of each other. Lattices play an important role for a multi-terminal source coding problem through a relay. Lattices have not played a central role in each of conventional multi-terminal source coding or point-to-point relay source coding problems in isolation. However, when the two problems are combined into one setting, we find that a lattice encoding and decoding framework is important for deriving non-trivial achievable rates for the system. This bears similarity to the distributed function computation and other computational coding contexts where again lattices perform well when compared to non-lattice based schemes.

We presented a lattice coding scheme for the distributed source channel communication of linear functions of two jointly Gaussian sources. We showed

that for any general linear function, we can find the scaling parameter γ to achieve a distortion very close to the lower bound (within a constant gap of the optimum) if $\mathrm{SNR} > \frac{1}{2}$. Moreover, the scheme achieves a smaller distortion than uncoded transmission for the difference function when $\rho > \frac{1}{2}(1+\frac{1}{2}\mathrm{SNR}^{-1})$. Future work includes considering scenarios with bandwidth mismatch and more transmitters.

We suspect that, regardless of the nature of the function to be computed (identity map or otherwise), lattice-based analysis will have a significant impact on multi-terminal source-coding problems over networks. In particular, lattices can aid the process of obtaining constant-gap inner and outer bounds for such problem settings. In summary, we believe the example networks studied in this dissertaion are just a couple of the many cases where lattices prove useful in distributed compression over general multi-hop networks.

We studied the problem of reduced reference image quality assessment by measuring the changes in suitably weighted entropies between the reference and distorted images in the wavelet domain. A distinguishing feature of the RRED indices is that these algorithms do not depend on any parameters that need to be trained on databases. The algorithms differ in the nature of the distortion measurement (by computing quality in different orientated subbands at different scales) and the quantity of the information required from the reference to compute quality (by filtering and subsampling in every subband). When the number of scalars required is around 1/40 of the image size, the algorithm achieves a performance which is nearly as good as the best perform-

ing full reference QA algorithms. Even when only a single scalar is obtained from the reference image, the algorithm achieves close to state of the art performance within each distortion category without knowing anything about the type of distortions that the image might have been subjected to. Moreover, the algorithms perform much better than mean squared error, which is an FR algorithm.

The overall performance of the single number algorithms may be improved further by better aligning the scores obtained for different distortion categories. This is a subject of future research. The use of a multiscale multiorientation decomposition before computing the index increases the complexity of the algorithm. Efficient implementations of this step can help reduce the time taken for implementation of the algorithm. The dependency on the viewing distance of the performance of such multiscale algorithms also requires a better understanding.

We also developed a family of RR VQA algorithms that vary in the amount of reference information required for quality computation. These algorithms are based on statistical models for video in both the spatial and temporal domains and compute the differences in the amount of information between the reference and the distorted videos to measure quality. While the algorithms with more information from the reference approach the performance of full reference VQA algorithms, the single number algorithm outperforms PSNR. Depending on the application and amount of reference information that can be afforded, a user could pick an algorithm from this class for

automatic prediction of video quality. The computation of the wavelet transform is a bottleneck in terms of the time complexity of the algorithm. The idea developed in this dissertation could be useful in other video representations with well behaved statistical models and faster computational times that will enable faster implementations of the algorithm.

Appendices

Appendix A

Proofs belonging to Chapter 2

A.1 Proof of Theorem 2.3

In order to prove Theorem 2.3, we need to show that $\forall \delta > 0$, there exist $(R'_{11}, R'_{12}, R'_{21}, R'_{22})$ and $(R_{11}, R_{12}, R_{21}, R_{22})$ that satisfy (2.12) and (2.13) such that

$$|R_{11}+R_{21}+R_{12}+R_{22}-I(X_1,X_2;U_{11},U_{12},U_{21},U_{22})-I(U_{11},U_{21};U_{12},U_{22})| \leq \delta.$$

Let $\epsilon = \frac{\delta}{8}$ and $(U_{11}, U_{12}, U_{21}, U_{22})$ satisfy the Markov condition $(U_{11}, U_{12}) - X_1 - X_2 - (U_{21}, U_{22})$. We choose

$$\begin{aligned}
R'_{11} &= I(X_1;U_{11}) + \epsilon & R'_{12} &= I(X_1;U_{12}|U_{11}) + I(U_{11};U_{12}) + \epsilon \\
R'_{21} &= I(X_2;U_{21}) + \epsilon & R'_{22} &= I(X_2;U_{22}|U_{21}) + I(U_{21};U_{22}) + \epsilon \\
R_{11} &= R'_{11} - I(U_{11};U_{21}) + \epsilon & R_{21} &= R'_{21} + \epsilon \\
R_{12} &= R'_{12} - I(U_{12};U_{22}) + \epsilon & R_{22} &= R'_{22} + \epsilon.
\end{aligned}$$

Note that $(R'_{11}, R'_{12}, R'_{21}, R'_{22})$ satisfy (2.12) and $(R_{11}, R_{12}, R_{21}, R_{22})$ satisfy (2.13). Therefore,

$$\begin{aligned}
R_{11}+R_{21}+R_{12}+R_{22} =& R'_{11}+R'_{12}+R'_{21}+R'_{22}-I(U_{11};U_{21})-I(U_{12};U_{22})+4\epsilon \\
=& I(X_1;U_{11},U_{12})+I(U_{11};U_{12})+I(X_2;U_{21},U_{22}) \\
&+I(U_{21};U_{22})-I(U_{11};U_{21})-I(U_{12};U_{22})+8\epsilon \\
=& I(X_1,X_2;U_{11},U_{12},U_{21},U_{22})+I(U_{11},U_{21};U_{12},U_{22})+\delta.
\end{aligned}$$

Allowing $\delta \to 0$, we see that we can achieve the sum rate

$$I(X_1,X_2;U_{11},U_{12},U_{21},U_{22})+I(U_{11},U_{21};U_{12},U_{22}),$$

for all auxiliaries $(U_{11},U_{12},U_{21},U_{22})$ such that $(U_{11},U_{12})-X_1-X_2-(U_{21},U_{22})$.

A.2 Proof of Lemma 2.11

By procedural steps, we have

$$\begin{aligned}
&I(X_1,X_2;U_{11},U_{21},U_{12},U_{22})+I(U_{11},U_{21};U_{12},U_{22}) \\
&\quad= I(S;U_{11},U_{21},U_{12},U_{22})+I(X_1,X_2;U_{11},U_{21},U_{12},U_{22}|S) \\
&\qquad+I(U_{11},U_{21};U_{12},U_{22}) \\
&\quad= I(S;U_{11},U_{21},U_{12},U_{22})+I(X_1;U_{11},U_{12}|S)+I(X_2;U_{21},U_{22}|S) \\
&\qquad+I(U_{11},U_{21};U_{12},U_{22}) \\
&\quad= I(S;U_{11},U_{21},U_{12},U_{22})+t'_1+t'_2+I(U_{11},U_{21};U_{12},U_{22}).
\end{aligned}$$

Recall that $Y_1 = X_1 + Z_1$ and $Y_2 = X_2 + Z_2$ where Z_1 and Z_2 are Gaussians with mean zero and variance $\sigma^2_{Z_1}$ and $\sigma^2_{Z_2}$ and independent of S, X_1 and X_2.

Now,

$$
\begin{aligned}
&I(U_{11},U_{21};U_{12},U_{22}) \\
&\quad = h(U_{11},U_{21}) + h(U_{12},U_{22}) - h(U_{11},U_{21},U_{12},U_{22}) \\
&\quad = h(U_{11},U_{21}) + h(U_{12},U_{22}) - h(U_{11},U_{21},U_{12},U_{22}) \\
&\qquad - h(U_{11},U_{21}|S,Y_1,Y_2) - h(U_{12},U_{22}|S,Y_1,Y_2) \\
&\qquad + h(U_{11},U_{21},U_{12},U_{22}|S,Y_1,Y_2) + I(U_{11},U_{21};U_{12},U_{22}|Y_1,Y_2,S) \\
&\quad = I(S,Y_1,Y_2;U_{11},U_{21}) + I(S,Y_1,Y_2;U_{12},U_{22}) \\
&\qquad - I(S,Y_1,Y_2;U_{11},U_{12},U_{21},U_{22}) + I(U_{11};U_{12}|S,Y_1) + I(U_{21};U_{22}|S,Y_2).
\end{aligned} \tag{A.1}
$$

For $l = 1, 2$, let $\delta_l = \sigma_S^2 e^{-2I(S;U_{1l},U_{2l})}$. Now, we can compute mutual information expressions between Gaussian random variables or use the fact that Gaussian random variables satisfy Lemma 3.1 in [51] with equality to conclude that,

$$
\begin{aligned}
\frac{1}{\sigma_S^2} e^{2I(S;U_{1l},U_{2l})} &= \frac{1}{\sigma_S^2} + \frac{1 - e^{-2I(X_1;U_{1l}|S)}}{\sigma_{N_1}^2} + \frac{1 - e^{-2I(X_2;U_{2l}|S)}}{\sigma_{N_2}^2} \\
\Rightarrow \frac{1}{\delta_l} &= \frac{1}{\sigma_S^2} + \frac{1}{\sigma_{N_1}^2} + \frac{1}{\sigma_{N_2}^2} - \frac{d'_{1l}}{\sigma_{N_1}^4} - \frac{d'_{2l}}{\sigma_{N_2}^4}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
I(S,Y_1,Y_2;U_{1l},U_{2l}) &= I(S;U_{1l},U_{2l}) + I(Y_1;U_{1l}|S) + I(Y_2;U_{2l}|S) \\
&= \frac{1}{2} \log \frac{\sigma_S^2(\sigma_{N_1}^2 + \sigma_{Z_1}^2)(\sigma_{N_2}^2 + \sigma_{Z_2}^2)}{\delta_l (d'_{1l} + \sigma_{Z_1}^2)(d'_{2l} + \sigma_{Z_2}^2)}.
\end{aligned} \tag{A.2}
$$

Observe that

$$\begin{aligned} &I(S, Y_1, Y_2; U_{11}, U_{12}, U_{21}, U_{22}) \\ &\quad = I(S; U_{11}, U_{21}, U_{12}, U_{22}) + I(Y_1, Y_2; U_{11}, U_{21}, U_{12}, U_{22}|S) \\ &\quad = I(S; U_{11}, U_{21}, U_{12}, U_{22}) + I(Y_1; U_{11}, U_{12}|S) + I(Y_2; U_{21}, U_{22}|S), \end{aligned} \tag{A.3}$$

and for $k = 1, 2$

$$\begin{aligned} I(Y_k; U_{k1}, U_{k2}|S) &= -h(Y_k|S, U_{k1}, U_{k2}) + h(Y_k|S) \\ &\overset{(a)}{=} -\frac{1}{2}\log(e^{\frac{2}{n}h(X_k|S,U_{k1},U_{k2})} + e^{\frac{2}{n}h(Z_k)}) + h(Y_k|S) \\ &= -\frac{1}{2}\log(e^{\frac{2}{n}(h(X_k|S)-I(X_k;U_{k1},U_{k2}|S))} + e^{\frac{2}{n}h(Z_k)}) + h(Y_k|S) \\ &= -\frac{1}{2}\log(\sigma_{N_k}^2 e^{-2t'_k} + \sigma_{Z_k}^2) + \frac{1}{2}\log(\sigma_{N_k}^2 + \sigma_{Z_k}^2), \end{aligned} \tag{A.4}$$

where (a) follows from EPI for Gaussians. From (A.1), (A.2), (A.3) and (A.4),

$$\begin{aligned} I(U_{11}, U_{21}; U_{12}, U_{22}) =& \frac{1}{2}\log\frac{(\sigma_{N_1}^2 + \sigma_{Z_1}^2)(\sigma_{N_2}^2 + \sigma_{Z_2}^2)\sigma_S^4}{(d'_{12} + \sigma_{Z_1}^2)(d'_{22} + \sigma_{Z_2}^2)(d'_{11} + \sigma_{Z_1}^2)(d'_{21} + \sigma_{Z_2}^2)\delta_1\delta_2} \\ &+ \frac{1}{2}\log(\sigma_{N_1}^2 e^{-2t'_1} + \sigma_{Z_1}^2)(\sigma_{N_2}^2 e^{-2t'_2} + \sigma_{Z_2}^2) \\ &+ I(U_{11}; U_{12}|S, Y_1) + I(U_{21}; U_{22}|S, Y_2) \\ &- I(S; U_{11}, U_{21}, U_{12}, U_{22}) \end{aligned}$$

and

$$\begin{aligned} &I(X_1, X_2; U_{11}, U_{12}, U_{21}, U_{22}) + I(U_{11}, U_{21}; U_{12}, U_{22}) \\ &\quad = r_1(d'_{11}, d'_{12}, t'_1, \sigma_{Z_1}^2) + r_2(d'_{21}, d'_{22}, t'_2, \sigma_{Z_2}^2) + \frac{1}{2}\log\frac{\sigma_S^4}{\delta_1\delta_2} \\ &\qquad + I(U_{11}; U_{12}|S, Y_1) + I(U_{21}; U_{22}|S, Y_2). \end{aligned}$$

Bibliography

[1] R. Ahlswede. The rate-distortion region for multiple descriptions without excess rate. *IEEE Trans. Inf. Theory*, 31:721–726, Nov. 1985.

[2] E. Ahmed and A. B. Wagner. Binary erasure multiple descriptions: Average-case distortion. In *Proc. Information Theory Workshop*, pages 166–170, 2009.

[3] E. Ahmed and A. B. Wagner. Binary erasure multiple descriptions: Worst-case distortion. In *Proc. IEEE Int Symp Info Theory*, pages 55–59, 2009.

[4] E. Ahmed and A. B. Wagner. Erasure multiple descriptions. *IEEE Trans. Inf. Theory*, to appear.

[5] S. H. Bae, T. N. Pappas, and B. Juang. Subjective evaluation of spatial resolution and quantization noise tradeoffs. *IEEE Trans. Image Process.*, 18:495–508, Mar. 2009.

[6] M. Barkowsky, J. Bialkowski, B. Eskoer, R. Bitto, and A. Kaup. Temporal trajectory aware video quality measure. *IEEE J. Sel. Topics Signal Process.*, 3(2):266–279, Apr. 2009.

[7] D. C. Burr and J. Ross. Contrast sensitivity at high velocities. *Vision Res.*, 23:3567–3569, 1982.

[8] M. Carnec, P. Le Callet, and D. Barba. Objective quality assessment of color images based on a generic perceptual reduced reference. *Signal Processing: Image Communication*, 23(4):239–256, Apr. 2008.

[9] D. M. Chandler and S. S. Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Trans. Image Process.*, 16(9):2284–2298, Sept. 2007.

[10] J. Chen. Rate region of Gaussian multiple description coding with individual and central distortion constraints. *IEEE Trans. Inf. Theory*, 55:3991–4005, Sept. 2009.

[11] J. Chen and T. Berger. Robust distributed source coding. *IEEE Trans. Inf. Theory*, 54:3385–3398, Aug. 2008.

[12] K. Chono, Y. C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod. Reduced-reference image quality assessment using distributed source coding. In *Proc. IEEE International Conference on Multimedia and Expo*, Hannover, Germany, 2008.

[13] T. Cover and A. El Gamal. Achievable rates for multiple descriptions. *IEEE Trans. Inf. Theory*, 28:851–857, Nov. 1982.

[14] T. Cover, A. El Gamal, and M. Salehi. Multiple access channels with arbitrarily correlated sources. *IEEE Trans. Inf. Theory*, 26:648–657, Nov. 1980.

[15] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, 1999.

[16] P. Cuff, H. Su, and A. El Gamal. Cascade multiterminal source coding. In *Proc. IEEE Int Symp Info Theory*, Seoul, Korea, 2009.

[17] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Amer. A (Optics, Image Science and Vision)*, 2(7):1160–1169, 1985.

[18] S. Diggavi and T. M. Cover. Worst additive noise under covariance constraints. *IEEE Trans. Inf. Theory*, 47:3072–3081, Nov. 2001.

[19] A. El Gamal and Y. H. Kim. *Lecture notes on information theory*. 2010, available at http://arxiv.org/abs/1001.3404.

[20] U. Engelke, M. Kusuma, H. J. Zepernick, and M. Caldera. Reduced-reference metric design for objective perceptual quality assessment in wireless imaging. *Signal Processing: Image Communication*, 24(7):525–547, Aug. 2009.

[21] U. Erez, S. Litsyn, and R. Zamir. Lattices which are good for (almost) everything. *IEEE Trans. Inf. Theory*, 51:3401–3416, Oct. 2005.

[22] U. Erez and R. Zamir. Achieving $\frac{1}{2}\log(1+SNR)$ on the AWGN channel with lattice encoding and decoding. *IEEE Trans. Inf. Theory*, 50:2293–2314, Oct. 2004.

[23] J. Foley. Human luminance pattern-vision mechanisms: masking experiments require a new model. *J. Opt. Soc. Amer. A (Optics and Image Science)*, 11(6):1710–1719, June 1994.

[24] X. Gao, W. Lu, D. Tao, and X. Li. Image quality assessment based on multiscale geometric analysis. *IEEE Transactions on Image Processing*, 18(7):1409–1423, July 2009.

[25] M. Gastpar. Uncoded transmission is exactly optimal for a simple Gaussian sensor network. *IEEE Trans. Inf. Theory*, 54:5247–5251, Nov. 2008.

[26] M. D. Gaubatz and S. S. Hemami. On the nearly scale-independent rank behavior of image quality metrics. In *Proc. IEEE Int. Conf. on Image Processing*, San Diego, CA 2008.

[27] T. J. Goblick. Theoretical limitations on the transmission of data from analog sources. *IEEE Trans. Inf. Theory*, 11:558–567, Oct. 1965.

[28] I. P. Gunawan and M. Ghanbari. Reduced-reference video quality assessment using discriminative harmonic strength with motion consideration. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(1):71–83, Jan. 2008.

[29] D. B. Hamilton, D. G. Albrecht, and W. S. Geisler. Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. *Vision Res.*, 29:1285–1308, 1989.

[30] A. P. Hekstra, J. G. Beerends, D. Ledermann, F. E. de Caluwe, S. Kohler, R. H. Koenen, S. Rihs, M. Ehrsam, and D. Schlauss. PVQM - a perceptual video quality measure. *Signal Processing: Image Communication*, 17:781–798, 2002.

[31] Y. Kochman and R. Zamir. Joint CEO/MAC and relaying. In *46th Annual Allerton Conference on Communication, Control and Computation*, Monticello, IL 2008.

[32] Y. Kochman and R. Zamir. Joint Wyner-Ziv/Dirty-Paper Coding by Modulo-Lattice Modulation. *IEEE Trans. Inf. Theory*, 2008, submitted for publication. Preprint available at http://www.eng.tau.ac.il/ zamir/publications.html.

[33] D. Krithivasan and S. Pradhan. Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function. *IEEE Trans. Inf. Theory*, 55:5628–5651, Dec. 2009.

[34] A. Lapidoth and S. Tinguely. Sending a bi-variate Gaussian source over a Gaussian MAC. In *Proc. IEEE Int Symp Info Theory*, Seattle, WA 2006.

[35] A. Lapidoth and S. Tinguely. Sending a bi-variate Gaussian source over a Gaussian MAC. 2010, submitted for publication. Preprint available at http://arxiv.org/abs/0901.3314.

[36] G. Legge and J. Foley. Contrast masking in human vision. *J. Opt. Soc. Amer.*, 70(12):1458–1471, Dec. 1980.

[37] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing: Special issue on Visual Media Quality Assessment*, 3:202–211, Apr. 2009.

[38] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. Perceptual blur and ringing metrics: Application to JPEG2000. *Signal Processing: Image Communication*, 19(2):163–172, 2004.

[39] M. Masry, S. S. Hemami, and Y. Sermadevi. A scalable wavelet-based video distortion metric and applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(2):260–273, Feb. 2006.

[40] A. K. Moorthy and A. C. Bovik. Blind image quality assessment: From scene statistics to perceptual quality. *IEEE Trans. Image Process.*, 20(12), Dec. 2011.

[41] B. Nazer and M. Gastpar. Structured Random Codes and Sensor Network Coding Theorems. In *Proceedings of the 20th Biennial International Zurich Seminar on Commununication (IZS 2008)*, Zurich, Switzerland, 2008.

[42] B. Nazer and M. Gastpar. Computation over multiple-access channels. *IEEE Trans. Inf. Theory*, 53:3498–3516, Oct. 2007.

[43] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba. Considering temporal variations of spatial visual distortions in video quality assessment. *IEEE J. Sel. Topics Signal Process.*, 3(2):253–265, Apr. 2009.

[44] I. Ohzawa, G. Sclar, and R. D. Freeman. Contrast gain control in the cat visual cortex. *Nature*, 298(5871):266–268, July 1982.

[45] Y. Oohama. The rate distortion function for the quadratic Gaussian CEO problem. *IEEE Trans. Inf. Theory*, 44:55–67, May 1998.

[46] L. Ozarow. On a source coding problem with two channels and three receivers. *Bell Syst. Tech. J.*, 59:1909–1921, Dec. 1980.

[47] H. Permuter and T. Weissman. Cascade and triangular source coding with side information at the first two nodes. 2010, submitted for publication. Preprint available at http://arxiv.org/abs/0901.3314.

[48] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *IEEE Trans. Broadcast*, 50(3):312–322, Sept. 2004.

[49] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. TID2008 - A database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 10:30–45, 2009.

[50] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.*, 12:1338–1351, Nov. 2003.

[51] V. Prabhakaran, K. Ramachandran, and D. Tse. Rate region of the quadratic Gaussian CEO problem. In *Proc. IEEE Int Symp Info Theory*, Chicago, USA, 2004.

[52] J. A. Redi, P. Gastaldo, I. Heynderickx, and R. Zunino. Color distribution information for the reduced-reference assessment of perceived image quality. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(12):1757–1769, Dec. 2010.

[53] M. Rohani, A. N. Avanaki, S. Nader-Esfahani, and M. Bashirpour. A reduced reference video quality assessment method based on the human motion perception. In *Fifth International Symposium on Telecommunications (IST)*, 2010.

[54] S. Roth and M. Black. On the spatial statistics of optical flow. *Int. J. of Computer Vision*, 74(1):33–50, Aug. 2007.

[55] M. A. Saad, A. C. Bovik, and C. Charrier. Mode-based blind image quality assessment using natural DCT statistics. *IEEE Trans. Image Process.*, 2011, to appear.

[56] Sarnoff Corporation. JNDmetrix technology. available at http://www.sarnoff.com/products services/video vision/ jndmetrix/

downloads.asp.

[57] O. Schwartz and E. P. Simoncelli. Natural signal statistics and sensory gain control. *Nat. Neurosci.*, 4(8):819–825, Aug. 2001.

[58] S. Servetto. Lattice quantization with side information. In *Proc. of Data Compression Conference*, Snowbird, UT 2000.

[59] K. Seshadrinathan and A. C. Bovik. Multi-scale and scalable video quality assessment. In *Proc. IEEE Int Conf on Consumer Electronics*, Las Vegas, NV, 2008.

[60] K. Seshadrinathan and A. C. Bovik. Unifying analysis of full reference image quality assessment. In *Proc. IEEE Int Conf on Image Processing*, San Diego, CA, 2008.

[61] K. Seshadrinathan and A. C. Bovik. Motion-tuned spatio-temporal quality assessment of natural videos. *IEEE Trans. Image Process.*, 19:335–350, Feb. 2010.

[62] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack. Study of subjective and objective quality assessment of video. *IEEE Trans. Image Process.*, 19:1427–1441, June 2010.

[63] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Trans. Image Process.*, 15(2):430–444, Feb. 2006.

[64] H. R. Sheikh, A. C. Bovik, and L. K. Cormack. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Trans. Image Process.*, 14(11):1918–1927, Nov. 2005.

[65] H. R. Sheikh, A. C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.*, 14(12):2117–2128, Dec. 2005.

[66] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Live image quality assessment database release 2. available at http://live.ece.utexas.edu/research/quality.

[67] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Proc. IEEE Int Conf on Image Processing*, 1995.

[68] E. P. Simoncelli and D. J. Heeger. A model of neuronal responses in visual area MT. *Vision Res.*, 38(5):743–761, Mar. 1998.

[69] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19:471–480, July 1973.

[70] R. Soundararajan and A. C. Bovik. RRED indices: Reduced reference entropic differencing for image quality assessment. *IEEE Trans. Image Process.*, 21(2):517–526, Feb. 2012.

[71] R. Soundararajan and S. Vishwanath. Communicating the difference of correlated Gaussian sources over a MAC. In *Proc. of Data Compression Conference*, Snowbird, UT 2009.

[72] R. Soundararajan and S. Vishwanath. Hybrid coding for Gaussian broadcast channels with Gaussian sources. In *Proc. IEEE Int Symp Info Theory*, Seoul, Korea 2009.

[73] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *Proc. IEEE Conf. on Computer Vision and Pattern Recog.*, San Francisco, CA 2010.

[74] P. C. Teo and D. J. Heeger. Perceptual image distortion. *Proc. SPIE*, 2179:127–141, 1994.

[75] P. Thompson. Perceived rate of movement depends on contrast. *Vision Res.*, 22:377–380, 1982.

[76] P. Thompson, K. Brooks, and S. T. Hammett. Speed can go up as well as down at low contrast: Implications for models of motion perception. *Vision Res.*, 46:782–786, 2006.

[77] S. Y. Tung. *Multiterminal Source Coding*. PhD thesis, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, 1978.

[78] D. Vasudevan, C. Tian, and S. N. Diggavi. Lossy source coding for a cascade communication system with side-informations. In *44th An-*

nual Allerton Conference on Communication, Control and Computation, Monticello, IL 2006.

[79] H. Viswanathan and T. Berger. The quadratic Gaussian CEO problem. *IEEE Trans. Inf. Theory*, 43:1549–1559, Sept. 1997.

[80] A. Wagner. On distributed compression of linear functions. In *46th Annual Allerton Conference on Communication, Control and Computation*, Monticello, IL 2008.

[81] A. B. Wagner. An outer bound for distributed compression of linear functions. In *42nd Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ 2008.

[82] A. B. Wagner. On distributed compression of linear functions. *IEEE Trans. Inf. Theory*, 57:79–94, Jan. 2011.

[83] A. B. Wagner, S. Tavildar, and P. Viswanath. Rate region of the Quadratic Gaussian Two-Encoder Source-Coding Problem. *IEEE Trans. Inf. Theory*, 54:1938–1961, May 2008.

[84] M. J. Wainwright and E. P. Simoncelli. Scale mixtures of Gaussians and the statistics of natural images. In *Adv. Neural Information Processing Systems*, volume 12, pages 64–68, May 2000.

[85] H. Wang and P. Viswanath. Vector Gaussian multiple descriptions with individual and centralized receivers. *IEEE Trans. Inf. Theory*, 53:2133–2153, June 2007.

[86] H. Wang and P. Viswanath. Vector Gaussian multiple description with two levels of receivers. *IEEE Trans. Inf. Theory*, 55:401–410, Jan. 2009.

[87] J. Wang, J. Chen, and X. Wu. On the minimum sum rate of Gaussian multiterminal source coding: new proofs. In *Proc. IEEE Int Symp Info Theory*, Seoul, Korea, 2009.

[88] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? - A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, Jan. 2009.

[89] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error measurement to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, Apr. 2004.

[90] Z. Wang and Q. Li. Video quality assessment using a statistical model of human visual speed perception. *J. Opt. Soc. Amer. A (Optics, Image Science and Vision)*, 24(12):B61–B69, Dec. 2007.

[91] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of JPEG compressed images. In *Proc. IEEE Int Conf on Image Processing*, Rochester, NY, 2002.

[92] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Proc. 37th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA 2003.

[93] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. H. Yang, and A. C. Bovik. Quality-aware images. *IEEE Trans. Image Process.*, 15(5):1680–1689, June 2006.

[94] A. Watson and J. Solomon. Model of visual contrast gain control and pattern masking. *J. Opt. Soc. Amer. A (Optics, Image Science and Vision)*, 14(9):2379–2391, Sept. 1997.

[95] A. B. Watson, J. Hu, and J. F. McGowan III. Digital video quality metric based on human vision. *Journal Electronic Imaging*, 10(1):20–29, Jan. 2001.

[96] M. P. Wilson, K. Narayanan, H. Pfister, and A. Sprintson. Joint physical layer and network coding for bi-directional relaying. *IEEE Trans. Inf. Theory*, 56:5641–5654, Nov. 2010.

[97] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, 22:1–10, Jan. 1976.

[98] W. Xue and X. Mou. Reduced reference image quality assessment based on Weibull statistics. In *Second International Workshop on Quality of Multimedia Experience*, Trondheim, 2010.

[99] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao. Two-terminal video coding. *IEEE Trans. Image Process.*, 18:534–551, Mar. 2009.

[100] R. Zamir. Lattices are everywhere. In *Proc. 2009 ITA Workshop*, San Diego, CA 2009.

[101] R. Zamir, S. Shamai, and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Inf. Theory*, 48:1250–1276, Nov. 2002.

Vita

Rajiv Soundararajan received the B. E. (Hons.) degree in electrical and electronics engineering from the Birla Institute of Technology and Science (BITS), Pilani, India, in 2006 and the M. S. degree in electrical and computer engineering from The University of Texas at Austin in 2008. He is currently pursuing the Ph. D. degree at The University of Texas at Austin. His research interests include statistical signal processing, information theory, distributed compression and image and video quality assessment.

Permanent address: 14 Majestic Park
Muthukumarappa Road
Saligramam, Chennai, India - 600093

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.